A medical student is conducting a chart review of patients admitted to the local emergency department with acute pancreatitis. He is interested in comparing blood glucose levels between mild and severe cases. Before attempting to compare the mean blood glucose levels between the 2 groups of patients, he decides to conduct a descriptive analysis of the variables in each group. Blood glucose levels in patients with severe acute pancreatitis were found to have a strong positively skewed distribution. Which of the following is most likely to be true regarding the data for the severe acute pancreatitis group?

A. The mean is equal to the median

B. The mean is equal to the mode

C. The mean is greater than the median

D. The median is greater than the mean

E. The mode is greater than the mean

Submit

A medical student is conducting a chart review of patients admitted to the local emergency department with acute pancreatitis. He is interested in comparing blood glucose levels between mild and severe cases. Before attempting to compare the mean blood glucose levels between the 2 groups of patients, he decides to conduct a descriptive analysis of the variables in each group. Blood glucose levels in patients with severe acute pancreatitis were found to have a strong positively skewed distribution. Which of the following is most likely to be true regarding the data for the severe acute pancreatitis group?

A. The mean is equal to the median (0%)

B. The mean is equal to the mode (0%)

✓ C. The mean is greater than the median (76%)

D. The median is greater than the mean (14%)

E. The mode is greater than the mean (7%)

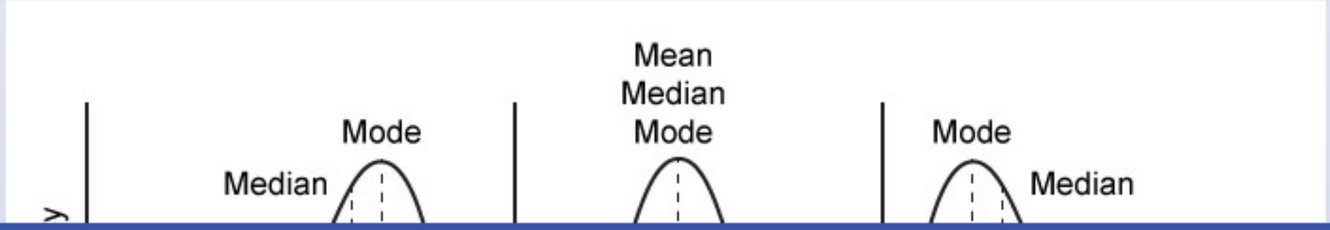
Omitted
Correct answer
C

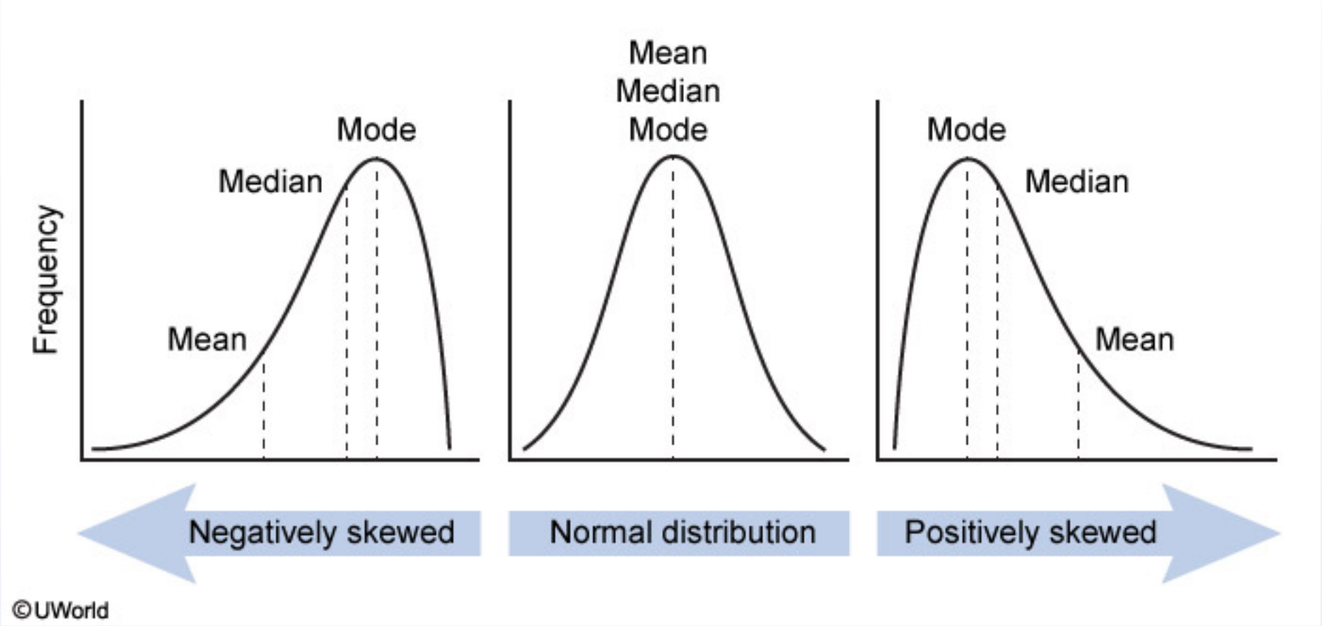
76%
Answered correctly

01 min, 58 secs
Time Spent

04/28/2020
Last Updated

Explanation



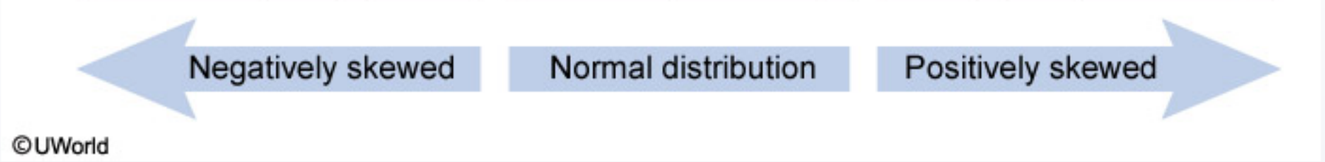


The symmetrical bell-shaped curve is usually cited as the prototypical curve describing the distribution of a continuous variable. However, in actual research scenarios, continuous variables often have **asymmetrical distributions**. When a distribution curve is asymmetric, it is either positively or negatively skewed:

- In a **positively skewed** distribution, there is an increased number of observations with **larger-than-expected magnitudes** (ie, extreme values) that shift the mean toward the **right**, producing a longer slope of the curve ("tail") on the positive side of the distribution.
- In a **negatively skewed** distribution, there is an increased number of observations with **smaller-than-expected magnitudes** that shift the mean toward the **left**, producing a longer slope of the curve ("tail") on the negative side of the distribution.

In general, in a **positively** skewed distribution, the mean is the most shifted in the positive direction, followed by the median and then the mode (**mode < median < mean**). In such a situation, the median often reflects a central tendency better than the mean does.

(Choices A and B) In perfectly normal distribution curves where there is no skew, all 3 measures of central tendency are precisely equal: mean = median = mode. If a minor skew is present, the 3 measures are approximately equal.



The symmetrical bell-shaped curve is usually cited as the prototypical curve describing the distribution of a continuous variable. However, in actual research scenarios, continuous variables often have **asymmetrical distributions**. When a distribution curve is asymmetric, it is either positively or negatively skewed:

- In a **positively skewed** distribution, there is an increased number of observations with **larger-than-expected magnitudes** (ie, extreme values) that shift the mean toward the **right**, producing a longer slope of the curve ("tail") on the positive side of the distribution.
- In a **negatively skewed** distribution, there is an increased number of observations with **smaller-than-expected magnitudes** that shift the mean toward the **left**, producing a longer slope of the curve ("tail") on the negative side of the distribution.

In general, in a **positively** skewed distribution, the mean is the most shifted in the positive direction, followed by the median and then the mode (**mode < median < mean**). In such a situation, the median often reflects a central tendency better than the mean does.

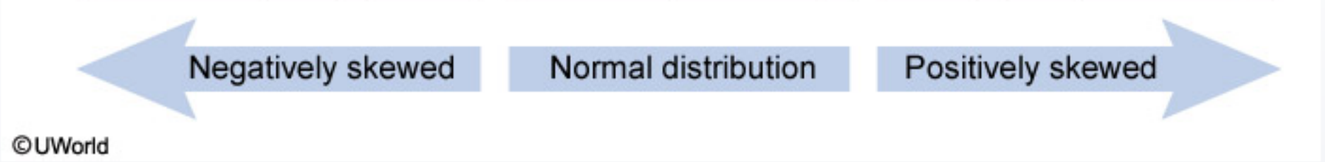
(Choices A and B) In perfectly normal distribution curves where there is no skew, all 3 measures of central tendency are precisely equal: mean = median = mode. If a minor skew is present, the 3 measures are approximately equal.

(Choices D and E) In general, in a negatively skewed distribution (with the "tail" on the left), the mean is the most shifted in the negative direction, followed by the median and then the mode.

Educational objective:
In general, in a positively skewed distribution, the mean is the most shifted in the positive direction (to the right), followed by the median and then the mode. In such a situation, the median often reflects a central tendency better than the mean does.

Biostatistics	Biostatistics & Epidemiology	Normal distribution
Subject	System	Topic

Copyright © UWorld. All rights reserved.



The symmetrical bell-shaped curve is usually cited as the prototypical curve describing the distribution of a continuous variable. However, in actual research scenarios, continuous variables often have **asymmetrical distributions**. When a distribution curve is asymmetric, it is either positively or negatively skewed:

- In a **positively skewed** distribution, there is an increased number of observations with **larger-than-expected magnitudes** (ie, extreme values) that shift the mean toward the **right**, producing a longer slope of the curve ("tail") on the positive side of the distribution.
- In a **negatively skewed** distribution, there is an increased number of observations with **smaller-than-expected magnitudes** that shift the mean toward the **left**, producing a longer slope of the curve ("tail") on the negative side of the distribution.

In general, in a **positively** skewed distribution, the mean is the most shifted in the positive direction, followed by the median and then the mode (**mode < median < mean**). In such a situation, the median often reflects a central tendency better than the mean does.

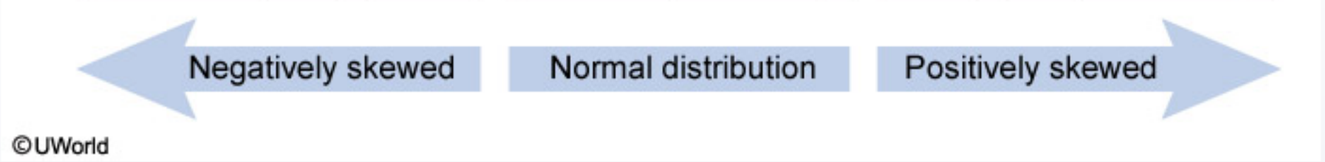
(Choices A and B) In perfectly normal distribution curves where there is no skew, all 3 measures of central tendency are precisely equal: mean = median = mode. If a minor skew is present, the 3 measures are approximately equal.

(Choices D and E) In general, in a negatively skewed distribution (with the "tail" on the left), the mean is the most shifted in the negative direction, followed by the median and then the mode.

Educational objective:
In general, in a positively skewed distribution, the mean is the most shifted in the positive direction (to the right), followed by the median and then the mode. In such a situation, the median often reflects a central tendency better than the mean does.

Biostatistics	Biostatistics & Epidemiology	Normal distribution
Subject	System	Topic

Copyright © UWorld. All rights reserved.



The symmetrical bell-shaped curve is usually cited as the prototypical curve describing the distribution of a continuous variable. However, in actual research scenarios, continuous variables often have **asymmetrical distributions**. When a distribution curve is asymmetric, it is either positively or negatively skewed:

- In a **positively skewed** distribution, there is an increased number of observations with **larger-than-expected magnitudes** (ie, extreme values) that shift the mean toward the **right**, producing a longer slope of the curve ("tail") on the positive side of the distribution.
- In a **negatively skewed** distribution, there is an increased number of observations with **smaller-than-expected magnitudes** that shift the mean toward the **left**, producing a longer slope of the curve ("tail") on the negative side of the distribution.

In general, in a **positively** skewed distribution, the mean is the most shifted in the positive direction, followed by the median and then the mode (**mode < median < mean**). In such a situation, the median often reflects a central tendency better than the mean does.

(Choices A and B) In perfectly normal distribution curves where there is no skew, all 3 measures of central tendency are precisely equal: mean = median = mode. If a minor skew is present, the 3 measures are approximately equal.

(Choices D and E) In general, in a negatively skewed distribution (with the "tail" on the left), the mean is the most shifted in the negative direction, followed by the median and then the mode.

Educational objective:
In general, in a positively skewed distribution, the mean is the most shifted in the positive direction (to the right), followed by the median and then the mode. In such a situation, the median often reflects a central tendency better than the mean does.

Biostatistics	Biostatistics & Epidemiology	Normal distribution
Subject	System	Topic

Copyright © UWorld. All rights reserved.

A study assessed the association between a new vaccine and traveler's diarrhea (TD). Researchers selected a random sample of people who intended to travel to regions where they were at increased risk for TD and who had received the new vaccine and another independent random sample of people who intended to travel to the same regions and who had not received the new vaccine. These 2 samples of travelers were assessed for the occurrence of TD during the trip and for 7 days after returning home. Which of the following measures of association are the investigators most likely to report?

A. Incidence

B. Median survival time

C. Odds ratio

D. Prevalence

E. Relative risk

Submit

A study assessed the association between a new vaccine and traveler's diarrhea (TD). Researchers selected a random sample of people who intended to travel to regions where they were at increased risk for TD and who had received the new vaccine and another independent random sample of people who intended to travel to the same regions and who had not received the new vaccine. These 2 samples of travelers were assessed for the occurrence of TD during the trip and for 7 days after returning home. Which of the following measures of association are the investigators most likely to report?

A. Incidence (10%)

B. Median survival time (0%)

C. Odds ratio (20%)

D. Prevalence (1%)

✓ E. Relative risk (66%)

Omitted
Correct answer
E

66%
Answered correctly

02 secs
Time Spent

05/12/2020
Last Updated

Explanation

This case describes a cohort study in which participants are initially selected based on their **exposure status to a risk modifier**: exposed (eg, vaccinated) or nonexposed (eg, not vaccinated). Once participants are categorized based on exposure status, the occurrence (ie, incidence) of the outcome/disease of interest (ie, traveler's diarrhea [TD]) within a specific period is determined in each group. The **risk of disease** is finally compared between exposed and nonexposed groups to estimate the **association** between exposure to a risk modifier and disease occurrence. A common measure of association in cohort studies is the **relative risk (RR)**. The RR is calculated as follows:

Explanation

This case describes a cohort study in which participants are initially selected based on their **exposure status to a risk modifier**: exposed (eg, vaccinated) or nonexposed (eg, not vaccinated). Once participants are categorized based on exposure status, the occurrence (ie, incidence) of the outcome/disease of interest (ie, traveler's diarrhea [TD]) within a specific period is determined in each group. The **risk of disease** is finally compared between exposed and nonexposed groups to estimate the **association** between exposure to a risk modifier and disease occurrence. A common measure of association in cohort studies is the **relative risk (RR)**. The RR is calculated as follows:

RR = (risk of disease in exposed) / (risk of disease in nonexposed)

The exposed group could also have been defined as the unvaccinated group and the nonexposed group as the vaccinated group; in this case, the RR would be inversed (interpretation of RR depends on what is chosen to represent the exposure).

(Choices A and D) Incidence (ie, new occurrences of a disease in a population over a period) and prevalence (ie, existing cases of a disease in a population at a point in time) are measures of disease burden. They do not directly convey information regarding the relationship between exposure and disease.

(Choice B) Median survival time is calculated in cohort studies or clinical trials and is used to compare the median time to an event of interest (eg, TD) in ≥2 groups of individuals (eg, those not vaccinated and those vaccinated). The purpose of this study is to assess the association between exposure to a risk modifier (eg, vaccination) and disease occurrence (eg, TD), not to compare the median time to develop the disease (eg, TD) between exposed and nonexposed groups.

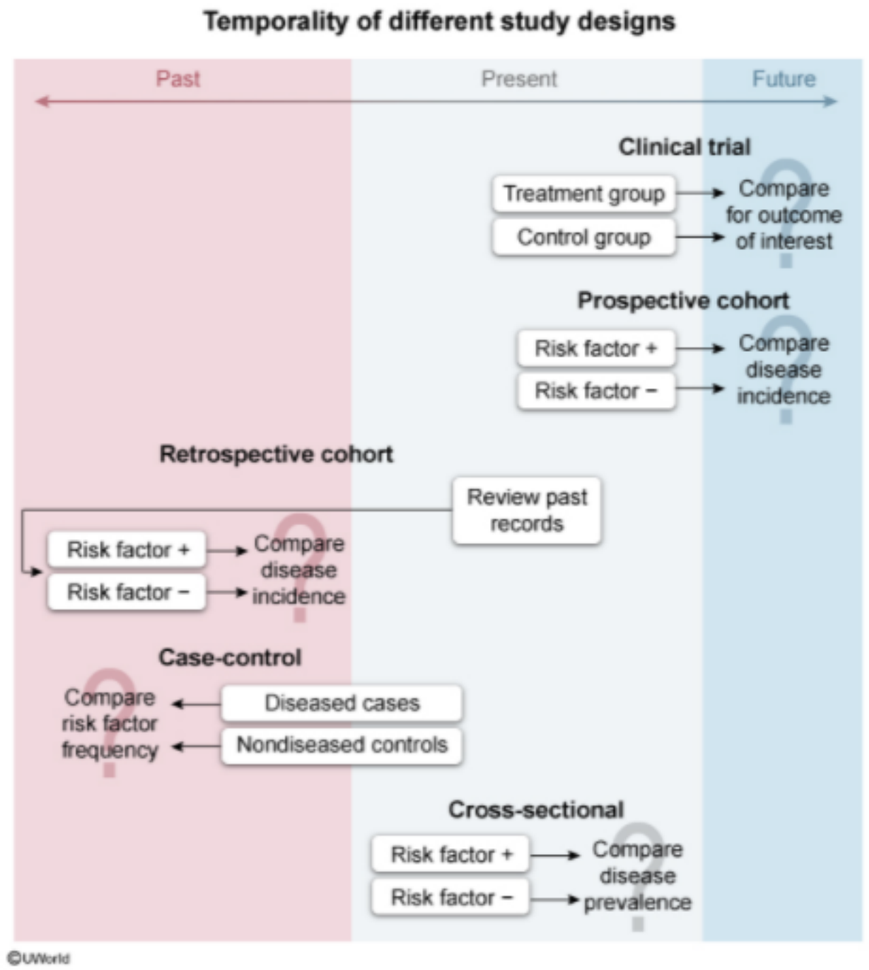
(Choice C) Odds ratio is calculated in case-control studies, where 2 groups of individuals are initially identified as "cases" or "controls" according to an outcome (eg, disease of interest), and then the frequency of exposure to the risk factors is compared between cases and controls to estimate the association between the risk factors and the outcomes.

Educational objective:
A cohort study is used to compare incidence of disease between exposed and nonexposed individuals. A common measure of association in cohort studies is the relative risk.

Biostatistics | Biostatistics & Epidemiology | Measures of association

Explanation

Exhibit Display



Zoom In Zoom Out Reset Add To Flash Card

A new test to diagnose urinary tract infections in women is being evaluated. The comparison gold standard is positive urine dipstick plus urine culture. Results of the study are given below.

	Positive urine dipstick & culture	Negative urine dipstick & culture
Test positive	60	20
Test negative	140	180

What is the new test's specificity?

A. 10%
B. 30%
C. 50%
D. 70%
E. 90%

Submit

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

A new test to diagnose urinary tract infections in women is being evaluated. The comparison gold standard is positive urine dipstick plus urine culture. Results of the study are given below.

	Positive urine dipstick & culture	Negative urine dipstick & culture
Test positive	60	20
Test negative	140	180

What is the new test's specificity?

A. 10% (5%)
B. 30% (8%)
C. 50% (3%)
D. 70% (3%)
✓ E. 90% (79%)

Omitted
Correct answer
E

79%
Answered correctly

05 secs
Time Spent

02/14/2020
Last Updated

Explanation

Positive condition	Negative condition

Explanation

	Positive condition	Negative condition	
Positive test result	TP	FP	PPV = TP/(TP + FP)
Negative test result	FN	TN	NPV = TN/(TN + FN)
	Sensitivity = TP/(TP + FN)	Specificity = TN/(TN + FP)	

FN = false negative; **FP** = false positive; **NPV** = negative predictive value; **PPV** = positive predictive value; **TN** = true negative; **TP** = true positive.

The **specificity** of a test refers to its ability to correctly identify individuals without the disease; it is the number of true negatives (TN) divided by all those without the disease. Specificity should be high in **confirmatory** tests to decrease the number of false positives (FP). Using a generic 2×2 (contingency) table, specificity can be calculated from: TN / (TN + FP). In this case:

Specificity = TN / (TN + FP) = 180 / (180 + 20) = 180/200 = 0.90 (or 90%)

(Choice A) In this example, FP / (TN + FP) = 20 / (180 + 20) = 20/200 = 10%. This value is called the false positive rate, which is equal to (1 – specificity). It represents the probability of a false positive test result in a patient who in reality is free of disease.

(Choices B and D) Sensitivity (a test's ability to correctly identify individuals with the disease) can be given by TP / (TP + FN), where TP is true positives and FN is false negatives. In this example, sensitivity = TP / (TP + FN) = 60 / (60 + 140) = 60/200 = 0.30 (or 30%). The false negative rate (= 1 – sensitivity) is given by FN / (TP + FN) = 140 / (60 + 140) = 140/200 = 0.70 (or 70%). It is the probability of a false negative test result in a patient who in reality has the disease.

(Choice C) From a total of 400 patients (= 60 + 140 + 20 + 180), there are 200 patients (= 60 + 140) who have a urinary tract infection (UTI)

Sensitivity = TP/(TP + FN)	Specificity = TN/(TN + FP)

FN = false negative; **FP** = false positive; **NPV** = negative predictive value; **PPV** = positive predictive value; **TN** = true negative; **TP** = true positive.

The **specificity** of a test refers to its ability to correctly identify individuals without the disease; it is the number of true negatives (TN) divided by all those without the disease. Specificity should be high in **confirmatory** tests to decrease the number of false positives (FP). Using a generic 2×2 (contingency) table, specificity can be calculated from: TN / (TN + FP). In this case:

Specificity = TN / (TN + FP) = 180 / (180 + 20) = 180/200 = 0.90 (or 90%)

(Choice A) In this example, FP / (TN + FP) = 20 / (180 + 20) = 20/200 = 10%. This value is called the false positive rate, which is equal to (1 – specificity). It represents the probability of a false positive test result in a patient who in reality is free of disease.

(Choices B and D) Sensitivity (a test's ability to correctly identify individuals with the disease) can be given by TP / (TP + FN), where TP is true positives and FN is false negatives. In this example, sensitivity = TP / (TP + FN) = 60 / (60 + 140) = 60/200 = 0.30 (or 30%). The false negative rate (= 1 – sensitivity) is given by FN / (TP + FN) = 140 / (60 + 140) = 140/200 = 0.70 (or 70%). It is the probability of a false negative test result in a patient who in reality has the disease.

(Choice C) From a total of 400 patients (= 60 + 140 + 20 + 180), there are 200 patients (= 60 + 140) who have a urinary tract infection (UTI) based on the gold standard. The prevalence of UTIs is therefore 200/400 = 0.50 (or 50%).

Educational objective:
The specificity of a test is its ability to correctly identify individuals without the disease. Specificity should be high in confirmatory tests to decrease false positives.

Biostatistics	Biostatistics & Epidemiology	Sensitivity, specificity, positive predictive value PPV, negative predictive value NPV
Subject	System	Topic

Copyright © UWorld. All rights reserved.

Drugs A and B are 2 new experimental drugs being tested for the treatment of a novel respiratory viral infection that causes acute respiratory failure and death in children. Part of the testing process is to analyze the 2-week survival after treatment to determine the clinical efficacy of the experimental drugs. A total of 60 children recently diagnosed with the disease are randomly assigned in a 1:1:1 ratio to receive Drug A, Drug B, or placebo. The absolute risk reduction of Drug A compared to placebo was found to be 0.05, whereas the absolute risk reduction of Drug B compared to placebo was found to be 0.20. Based on these results, which of the following statements comparing the effectiveness of Drugs A and B in treating children infected with the novel virus is most appropriate?

- A. Drugs A and B require treating the same number of children to prevent 1 additional death, so they are equally effective
- B. Drug A requires treating fewer children to prevent 1 additional death compared to Drug B, so Drug A is less effective than Drug B
- C. Drug A requires treating fewer children to prevent 1 additional death compared to Drug B, so Drug A is more effective than Drug B
- D. Drug A requires treating more children to prevent 1 additional death compared to Drug B, so Drug A is less effective than Drug B
- E. Drug A requires treating more children to prevent 1 additional death compared to Drug B, so Drug A is more effective than Drug B

Submit

Drugs A and B are 2 new experimental drugs being tested for the treatment of a novel respiratory viral infection that causes acute respiratory failure and death in children. Part of the testing process is to analyze the 2-week survival after treatment to determine the clinical efficacy of the experimental drugs. A total of 60 children recently diagnosed with the disease are randomly assigned in a 1:1:1 ratio to receive Drug A, Drug B, or placebo. The absolute risk reduction of Drug A compared to placebo was found to be 0.05, whereas the absolute risk reduction of Drug B compared to placebo was found to be 0.20. Based on these results, which of the following statements comparing the effectiveness of Drugs A and B in treating children infected with the novel virus is most appropriate?

A. Drugs A and B require treating the same number of children to prevent 1 additional death, so they are equally effective (0%)

B. Drug A requires treating fewer children to prevent 1 additional death compared to Drug B, so Drug A is less effective than Drug B (1%)

C. Drug A requires treating fewer children to prevent 1 additional death compared to Drug B, so Drug A is more effective than Drug B (11%)

✓ D. Drug A requires treating more children to prevent 1 additional death compared to Drug B, so Drug A is less effective than Drug B (84%)

E. Drug A requires treating more children to prevent 1 additional death compared to Drug B, so Drug A is more effective than Drug B (2%)

Omitted
Correct answer
D

84%
Answered correctly

02 secs
Time Spent

05/30/2020
Last Updated

Explanation

Common measures of therapeutic efficacy

Common measures of therapeutic efficacy		
Term	**Definition**	**Calculation**
Absolute risk reduction (ARR)	Percentage indicating the actual difference in event rate between control & treatment groups	ARR = control rate – treatment rate
Relative risk reduction (RRR)	Percentage indicating relative reduction in the treatment event rate compared to the control group	RRR = ARR/control rate
Relative risk (RR)	Ratio of the probability of an event occurring in the treatment group compared to the control group	RR = treatment rate/control rate
Number needed to treat (NNT)	Number of individuals that need to be treated to prevent a negative outcome in 1 patient	NNT = 1/ARR

The **absolute risk reduction** (ARR) describes the efficacy of a treatment (eg, Drug A) compared to a control group (eg, placebo); it is the difference in the risk (or rate) of a negative event (eg, death) between treatment and control groups:

$$ARR = (Risk_{control} - Risk_{treatment})$$

ARR expressed as a percentage describes the number of negative events (eg, deaths) prevented in 100 patients. For example, an ARR of 0.05 indicates that 5 of 100 patients treated with a treatment (eg, Drug A) would be prevented from developing a negative event. Therefore, to prevent 1 patient from developing a negative event, it would be necessary to treat 100 / 5 = 1 / 0.05 = 20 patients. This is the **number needed to treat (NNT)**.

NNT is the number of patients who need to be treated with a treatment (eg, Drug A) to **prevent 1 additional negative event** (eg, death) compared to a control group (eg, placebo); NNT is the **inverse of the ARR**:

ARR expressed as a percentage describes the number of negative events (eg, deaths) prevented in 100 patients. For example, an ARR of 0.05 indicates that 5 of 100 patients treated with a treatment (eg, Drug A) would be prevented from developing a negative event. Therefore, to prevent 1 patient from developing a negative event, it would be necessary to treat 100 / 5 = 1 / 0.05 = 20 patients. This is the **number needed to treat (NNT)**.

NNT is the number of patients who need to be treated with a treatment (eg, Drug A) to **prevent 1 additional negative event** (eg, death) compared to a control group (eg, placebo); NNT is the **inverse of the ARR**:

NNT = 1 / ARR

A **lower NNT** indicates **more effective treatments** because fewer patients would need to be treated to prevent 1 additional negative outcome. NNTs can be used to compare the effectiveness of different treatments within a single study or between similar studies (eg, similar patient characteristics, control groups, duration of follow-up). In this example, a single study is testing 2 treatments (ie, Drugs A and B) against placebo (ie, control group) to prevent death (ie, negative effect) in recently diagnosed children randomized to treatment groups with a similar follow-up (ie, 2-week survival):

- Drug A versus placebo: ARR = 0.05, so NNT = 1 / 0.05 = 20
- Drug B versus placebo: ARR = 0.20, so NNT = 1 / 0.20 = 5

Drug A requires treating more children to prevent 1 additional death compared to Drug B (20 vs 5). A higher NNT indicates a lower effectiveness; therefore, **Drug A is less effective** compared to Drug B **(Choices A, B, C, and E)**.

Educational objective:
The number needed to treat (NNT) is the number of patients who need to receive a treatment to prevent 1 additional negative event. NNT is the inverse of the absolute risk reduction. The lower the NNT, the more effective the treatment because fewer patients need be treated to prevent 1 additional negative event.

Biostatistics	Biostatistics & Epidemiology	Number Needed To Treat
Subject	System	Topic

Copyright © UWorld. All rights reserved.

In a reference sample of hundreds of healthy subjects, the laboratory reference range for a novel marker of cardiac injury is 0.04-0.08 U/mL at the standard 95% level of probability. The marker has very high sensitivity and specificity for myocardial tissue. The clinical cardiology team would like to use a 99.7% reference range to assess patients who come to the emergency department with chest pain and have a high pretest probability of cardiac ischemia. An elevated value of the marker is defined as exceeding the 99.7th percentile of the reference sample. Assuming a normal (Gaussian) distribution with a mean of 0.06 U/mL, which of the following most closely approximates the corresponding reference range?

A. 0.03 to 0.09

B. 0.035 to 0.085

C. 0.045 to 0.075

D. 0.05 to 0.07

E. 0.055 to 0.065

Submit

In a reference sample of hundreds of healthy subjects, the laboratory reference range for a novel marker of cardiac injury is 0.04-0.08 U/mL at the standard 95% level of probability. The marker has very high sensitivity and specificity for myocardial tissue. The clinical cardiology team would like to use a 99.7% reference range to assess patients who come to the emergency department with chest pain and have a high pretest probability of cardiac ischemia. An elevated value of the marker is defined as exceeding the 99.7th percentile of the reference sample. Assuming a normal (Gaussian) distribution with a mean of 0.06 U/mL, which of the following most closely approximates the corresponding reference range?

- ✓ A. 0.03 to 0.09 (73%)
- B. 0.035 to 0.085 (12%)
- C. 0.045 to 0.075 (5%)
- D. 0.05 to 0.07 (4%)
- E. 0.055 to 0.065 (3%)

Omitted
Correct answer
A

73%
Answered correctly

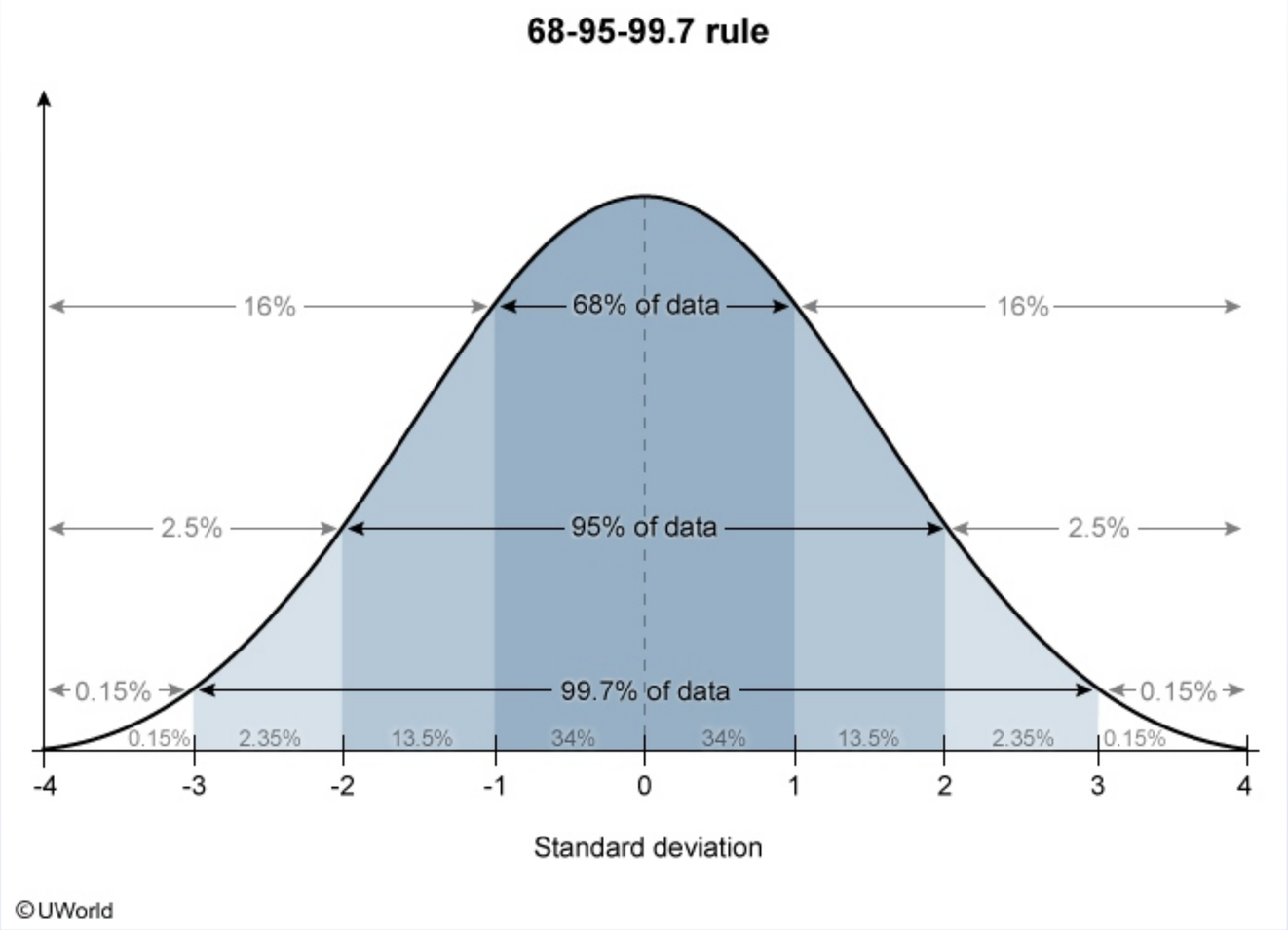
02 secs
Time Spent

06/03/2020
Last Updated

Explanation

68-95-99.7 rule

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31



A **normal (Gaussian) distribution** refers to a symmetrical, bell-shaped distribution with a fixed proportion of observations lying within a certain distance of the mean. This distance is called the **standard deviation (SD)** and reflects the **degree of dispersion** from the mean. According to the properties of this distribution, 68% of observations lie within 1 SD from the mean, with half (ie, 68/2 = 34%) above and half (34%) below the mean. The remaining 32% (= 100% − 68%) lie outside 1 SD from the mean, with half (ie, 32/2 = 16%) above and the other half (16%) below 1 SD from the mean. In addition, 95% of all observations lie within 2 SD of the mean, and 99.7% of all observations lie within 3 SD from the mean. This

A **normal (Gaussian) distribution** refers to a symmetrical, bell-shaped distribution with a fixed proportion of observations lying within a certain distance of the mean. This distance is called the **standard deviation (SD)** and reflects the **degree of dispersion** from the mean. According to the properties of this distribution, 68% of observations lie within 1 SD from the mean, with half (ie, 68/2 = 34%) above and half (34%) below the mean. The remaining 32% (= 100% − 68%) lie outside 1 SD from the mean, with half (ie, 32/2 = 16%) above and the other half (16%) below 1 SD from the mean. In addition, 95% of all observations lie within 2 SD of the mean, and 99.7% of all observations lie within 3 SD from the mean. This is the **68-95-99.7 rule**.

The 95% range for this marker with a **mean of 0.06** U/mL is 0.04 to 0.08 U/mL; given a normal distribution, this represents the range given by mean ± 2 SD. Therefore, **1 SD = 0.01** (because 0.04 = mean − 2 SD = 0.06 − 2 SD = 0.06 − 2 × 0.01 and, similarly, 0.08 = 0.06 + 2 × 0.01). The **99.7% range** is given by **mean ± 3 SD**; therefore, the 99.7% range is **0.03** (= mean − 3 SD = 0.06 − 3 × 0.01) to **0.09** (= mean + 3 SD = 0.06 + 3 × 0.01). Any result that falls outside this range (eg, a value exceeding this range) indicates that the value is different from what is seen in 99.7% of the reference population.

In laboratory measurements, although 95% is often used as a reference range, a 99% range is used in certain cases (eg, to determine cutoff for troponin in acute myocardial infarction). Other laboratory analysis methods involve performing more specific calculations when it is important to detect a difference in one direction (eg, for a cancer marker, knowing that the value is significantly higher than normal is more important than knowing that it is significantly different—either higher or lower—than normal). In some cases, high-sensitivity assays (eg, high-sensitivity C-reactive protein testing) allow the detection of values in the lower range of normal, which would not have been possible with standard assays.

Educational objective:
In a normal (bell-shaped) distribution, 68% of all values are within 1 standard deviation (SD) from the mean; 95% are within 2 SD from the mean; and 99.7% are within 3 SD from the mean.

References

- Statistical distributions commonly used in measurement uncertainty in laboratory medicine.
- Targeting residual inflammatory risk: a shifting paradigm for atherosclerotic disease.

Biostatistics	Biostatistics & Epidemiology	P-value and confidence interval
Subject	System	Topic

A meta-analysis of several trials on the effect of cocoa intake on systolic blood pressure (SBP) revealed the following results:

Study	Mean SBP in cocoa group – mean SBP in control group (mm Hg) [95% confidence interval]
1	−5.2 [−7.3, −3.4]
2	1.6 [−5.3, 10.4]
3	−4.2 [−8.1, −2.7]
4	−2.9 [−4.1, −1.4]
5	−2.8 [−5.2, −1.1]
6	0.8 [0.1, 1.2]
7	1.0 [−1.2, 3.3]
Total	**−2.2 [−2.7, −1.3]**

All the trials evaluated the difference in SBP at 2 weeks. Based on the data, which of the following is the most appropriate conclusion?

A. A higher mean SBP was seen in the cocoa groups overall

B. Cocoa intake should be recommended for blood pressure management

C. Cocoa intake was associated with a statistically significant decrease in SBP

D. Studies 2, 6, and 7 showed a statistically significant increase in SBP

E. There was no statistically significant change in SBP overall

Study	Mean SBP in cocoa group − mean SBP in control group (mm Hg) [95% confidence interval]
1	−5.2 [−7.3, −3.4]
2	1.6 [−5.3, 10.4]
3	−4.2 [−8.1, −2.7]
4	−2.9 [−4.1, −1.4]
5	−2.8 [−5.2, −1.1]
6	0.8 [0.1, 1.2]
7	1.0 [−1.2, 3.3]
Total	**−2.2 [−2.7, −1.3]**

All the trials evaluated the difference in SBP at 2 weeks. Based on the data, which of the following is the most appropriate conclusion?

A. A higher mean SBP was seen in the cocoa groups overall (2%)

B. Cocoa intake should be recommended for blood pressure management (2%)

✓ C. Cocoa intake was associated with a statistically significant decrease in SBP (71%)

D. Studies 2, 6, and 7 showed a statistically significant increase in SBP (5%)

E. There was no statistically significant change in SBP overall (18%)

Omitted
Correct answer

71%
Answered correctly

02 secs
Time Spent

03/20/2020
Last Updated

A **meta-analysis** groups the results of several trials (ideally, high-quality randomized controlled trials) to increase statistical power and provide an overall estimate of the effect of an exposure (eg, cocoa intake) on an outcome (eg, systolic blood pressure [SBP]). The individual results of the trials are assessed and weighted to obtain a **pooled effect estimate**. In this example, the meta-analysis showed an overall change in mean SBP of −2.2 mm Hg (ie, decrease in SBP) in the cocoa intake group compared to the control group. The 95% confidence interval (CI) was −2.7 to −1.3 mm Hg, indicating a statistically significant difference because the CI does not include the null value (explained below).

CIs give a range of plausible values for an unknown parameter (eg, difference between 2 mean SBPs) based on results from a sample. If the **CI does not include the null value**, then the result is **statistically significant**; if it crosses the null value, then the result is not statistically significant. All CIs have a null value, but the null value is not the same for all CIs. For an odds ratio or a relative risk (RR), the null value is 1 because these statistics are ratios (ie, RR = 1 represents no difference in risk between the groups). However, if the parameter of interest is a **difference** (eg, difference in mean SBP between cocoa intake and control groups), then the **null value is 0** because that represents no difference between the groups.

In this case, the 95% CI of [−2.7, −1.3] does not include the null value of 0 and therefore the result is **statistically significant**. In summary, cocoa intake was associated with a statistically significant decrease in SBP.

(Choices A and E) The overall mean change in SBP was negative at −2.2 mm Hg, reflecting a lower SBP in the cocoa group compared to the control group. The change was statistically significant.

(Choice B) Although there was a statistically significant decrease in SBP in the cocoa intake group, the clinical significance may be limited as the absolute SBP decrease is only about 2 mm Hg. Furthermore, the trials measured SBP at 2 weeks, so it is possible that the effect of cocoa intake on SBP was short-lived. Finally, no information is provided regarding adverse effects of cocoa intake.

(Choice D) Given that the CI crossed the null value in studies 2 and 7, there was no statistically significant difference between the groups in these studies.

Educational objective:
A meta-analysis groups results of several trials to increase statistical power and provide an overall pooled effect estimate.

Biostatistics Biostatistics & Epidemiology Meta analysis

A research laboratory is evaluating a new dipstick test for diagnosing urinary tract infections (UTIs). The new test is compared to urinalysis of a midstream urine specimen (considered the diagnostic gold standard) to establish diagnostic test parameters. The study enrolls 300 patients, of which 100 have a UTI as determined by urinalysis. The new dipstick test is determined to be 70% sensitive and 90% specific for the diagnosis of UTIs. How many false positives are present in the study?

- A. 20
- B. 30
- C. 70
- D. 120
- E. 180

Submit

A research laboratory is evaluating a new dipstick test for diagnosing urinary tract infections (UTIs). The new test is compared to urinalysis of a midstream urine specimen (considered the diagnostic gold standard) to establish diagnostic test parameters. The study enrolls 300 patients, of which 100 have a UTI as determined by urinalysis. The new dipstick test is determined to be 70% sensitive and 90% specific for the diagnosis of UTIs. How many false positives are present in the study?

- ✓ A. 20 (63%)
- B. 30 (25%)
- C. 70 (4%)
- D. 120 (2%)
- E. 180 (3%)

Omitted
Correct answer
A

63%
Answered correctly

02 secs
Time Spent

04/15/2020
Last Updated

Explanation

	Disease positive	Disease negative	
Test positive	a True positive	b False positive	a + b
Test negative	c False negative	d True negative	c + d

A research laboratory is evaluating a new dipstick test for diagnosing urinary tract infections (UTIs). The new test is compared to urinalysis of a midstream urine specimen (considered the diagnostic gold standard) to establish diagnostic test parameters. The study enrolls 300 patients, of which 100 have a UTI as determined by urinalysis. The new dipstick test is determined to be 70% sensitive and 90% specific for the diagnosis of UTIs. How many false positives are present in the study?

- ✓ A. 20 (63%)
- B. 30 (25%)
- C. 70 (4%)
- D. 120 (2%)
- E. 180 (3%)

Omitted
Correct answer
A

63%
Answered correctly

02 secs
Time Spent

04/15/2020
Last Updated

Explanation

	Disease positive	Disease negative	
Test positive	a True positive	b False positive	a + b
Test negative	c False negative	d True negative	c + d

	Disease positive	Disease negative	
Test positive	a True positive	b False positive	a + b
Test negative	c False negative	d True negative	c + d
	a + c	b + d	

Sensitivity = a/(a + c); Specificity = d/(b + d)

False positives describe people who **test positive** for a disease but **do not actually have it**.

The number of false positives can be calculated using the test's specificity and total number of patients without urinary tract infections (UTIs). The **specificity** of a test refers to the number of true negatives (d) divided by the total number of patients without the disease (b + d). In this case, the specificity is 90% (ie, 0.9) and the total number of patients without UTIs (b + d) equals 300 – 100 = 200. Using this information, the number of **true negatives** (d) can be calculated:

Specificity = d/(b + d)
0.9 = d/200
d = 180

The total number of patients without the disease (b + d) equals 200, and the number of true negatives (d) equals 180. Using this information, the number of false positives (b) can be calculated:

False positives = (b + d) − d = 200 – 180 = 20

For the sake of thoroughness, the 2 × 2 table can be completed as follows. **Sensitivity** refers to the number of true positives (a) divided by all people with the disease (a + c). Repeating the same exercise as above gives the completed table:

Sensitivity = a/(a + c)

The number of false positives can be calculated using the test's specificity and total number of patients without urinary tract infections (UTIs). The **specificity** of a test refers to the number of true negatives (d) divided by the total number of patients without the disease (b + d). In this case, the specificity is 90% (ie, 0.9) and the total number of patients without UTIs (b + d) equals 300 – 100 = 200. Using this information, the number of **true negatives** (d) can be calculated:

Specificity = d/(b + d)
0.9 = d/200
d = 180

The total number of patients without the disease (b + d) equals 200, and the number of true negatives (d) equals 180. Using this information, the number of false positives (b) can be calculated:

False positives = (b + d) − d = 200 – 180 = 20

For the sake of thoroughness, the 2 × 2 table can be completed as follows. **Sensitivity** refers to the number of true positives (a) divided by all people with the disease (a + c). Repeating the same exercise as above gives the completed table:

Sensitivity = a/(a + c)
0.7 = a/100
a = 70
c = (a + c) – a = 100 − 70 = 30

	+ UTI	- UTI	
Test positive	70	20	90
Test negative	30	180	210
	100	200	

The number of true negatives and false positives can be more quickly calculated using the following equations:

False positives = (b + d) − d = 200 – 180 = 20

For the sake of thoroughness, the 2 × 2 table can be completed as follows. **Sensitivity** refers to the number of true positives (a) divided by all people with the disease (a + c). Repeating the same exercise as above gives the completed table:

Sensitivity = a/(a + c)
0.7 = a/100
a = 70
c = (a + c) – a = 100 − 70 = 30

	+ UTI	- UTI	
Test positive	70	20	90
Test negative	30	180	210
	100	200	

The number of true negatives and false positives can be more quickly calculated using the following equations:

True negatives = (Specificity) * (Number of patients confirmed without the disease)
False positives = (1 − Specificity) * (Number of patients confirmed without the disease)

Educational objective:
Specificity is the number of true negatives divided by the total number of subjects confirmed as not having the disease.

True negatives = (Specificity) * (Number of patients confirmed without the disease)
False positives = (1 − Specificity) * (Number of patients confirmed without the disease)

Biostatistics — Subject
Biostatistics & Epidemiology — System
Sensitivity and specificity — Topic

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

A study determines that the mean blood cholesterol level is 195 mg/dL in 200 non-diabetic hospitalized patients and 210 mg/dL in 180 diabetic hospitalized patients. The probability that the observed difference is due to chance alone is reported to be 5%. There is also a 20% probability of concluding that there is no difference in blood cholesterol level when there is one in reality. What is the power of the study?

A. 0.05
B. 0.20
C. 0.50
D. 0.80
E. 0.95

Submit

A study determines that the mean blood cholesterol level is 195 mg/dL in 200 non-diabetic hospitalized patients and 210 mg/dL in 180 diabetic hospitalized patients. The probability that the observed difference is due to chance alone is reported to be 5%. There is also a 20% probability of concluding that there is no difference in blood cholesterol level when there is one in reality. What is the power of the study?

- A. 0.05 (7%)
- B. 0.20 (6%)
- C. 0.50 (3%)
- ✓ D. 0.80 (72%)
- E. 0.95 (10%)

Omitted
Correct answer
D

72%
Answered correctly

01 sec
Time Spent

07/04/2020
Last Updated

Explanation

The **power** of a study is the ability of a study to detect a difference between groups when such a difference truly exists. Power is related to **type II error (β)**, which is the probability of concluding there is no difference between groups when one truly exists. Mathematically, power is given by:

Power = 1 – β

In this example, the power of the study is the probability of detecting a difference in blood cholesterol level between diabetics and non-diabetics if there is a real difference. The probability of concluding that there is no difference in blood cholesterol level when in reality there is one is given as 20%; this corresponds to the definition of β (ie, β = 0.20 in this example). Therefore:

Correct answer
D

Answered correctly

Time Spent

Last Updated

Explanation

The **power** of a study is the ability of a study to detect a difference between groups when such a difference truly exists. Power is related to **type II error (β)**, which is the probability of concluding there is no difference between groups when one truly exists. Mathematically, power is given by:

Power = 1 – β

In this example, the power of the study is the probability of detecting a difference in blood cholesterol level between diabetics and non-diabetics if there is a real difference. The probability of concluding that there is no difference in blood cholesterol level when in reality there is one is given as 20%; this corresponds to the definition of β (ie, β = 0.20 in this example). Therefore:

Power = 1 – β = 1 – 0.20 = 0.80

(Choice A) Type I error (α) describes the probability of seeing a difference when there is no difference in reality. The value of α is generally compared to the probability that the observed difference is due to chance alone (a simplified explanation of the p-value). In this example, the probability that the observed difference between diabetic and nondiabetic patients is due to chance alone is given as 5% (0.05).

(Choices B and E) Type II error (β) is 0.20, as explained above. The value 0.95 corresponds to (1 – α), but power is given by (1 – β).

Educational objective:
The power of a study indicates the probability of seeing a difference when there is one. The formula is Power = 1 – β, where β is the type II error rate.

Biostatistics	Biostatistics & Epidemiology	Power and sample size
Subject	System	Topic

Copyright © UWorld. All rights reserved.

Public health officials involved in developing nutritional guidelines commission a study to determine how dietary eating patterns influence total body iron stores in children age 5-17. As part of the study, researchers want to assess how 2 independent variables, red meat consumption and egg consumption (both reported in units of ounces/week), affect serum ferritin concentrations while adjusting for age and gender. Which of the following statistical techniques is most helpful for determining the association between the study variables?

A. Analysis of variance

B. Meta-analysis

C. Odds ratio

D. Regression analysis

E. Relative risk

Submit

Public health officials involved in developing nutritional guidelines commission a study to determine how dietary eating patterns influence total body iron stores in children age 5-17. As part of the study, researchers want to assess how 2 independent variables, red meat consumption and egg consumption (both reported in units of ounces/week), affect serum ferritin concentrations while adjusting for age and gender. Which of the following statistical techniques is most helpful for determining the association between the study variables?

A. Analysis of variance (37%)

B. Meta-analysis (6%)

C. Odds ratio (5%)

✓ D. Regression analysis (40%)

E. Relative risk (9%)

Omitted
Correct answer
D

40%
Answered correctly

02 secs
Time Spent

05/15/2020
Last Updated

Explanation

		Dependent variable	
		Qualitative (categorical)	*Quantitative*
Independent variable	*Qualitative (categorical)*	Chi-square, logistic regression*	*t*-test, ANOVA, linear regression
	Quantitative	Logistic regression*	Correlation, linear regression

		Dependent variable	
		Qualitative (categorical)	*Quantitative*
Independent variable	*Qualitative (categorical)*	Chi-square, logistic regression*	*t*-test, ANOVA, linear regression
	Quantitative	Logistic regression*	Correlation, linear regression

* Dependent variable must be dichotomous.

ANOVA = analysis of variance.

Variables are broadly classified as qualitative (ie, categorical) or quantitative (eg, continuous) based on their scale of measurement. **Qualitative variables** (eg, type of treatment, blood type) **represent categories or groups** whereas **quantitative variables** (eg, temperature, glucose levels) **represent numerical values**. The scale of measurement of the dependent (eg, outcome) and independent (eg, exposures, risk factors) variables in a study determines the correct statistical test.

Regression analysis is used to describe the **association between 1 or more independent variables** (eg, exposures, risk factors), which may be quantitative or qualitative, and 1 quantitative dependent variable (ie, **a quantitative outcome** [eg, laboratory values]) while adjusting for other variables of interest. In this study:

- The quantitative dependent variable was serum ferritin concentration.
- The independent variables were consumption of red meat and eggs reported in ounces/week (2 quantitative variables).
- The effect of the main independent variables (eg, consumption of red meat and eggs) on the dependent variable (eg, serum ferritin concentration) is adjusted to account for the effects of other independent variables, which are called adjustment variables (eg, age, gender). Controlling for these variables reduces the potential confounding.

Calculation of a regression line (eg, line of best fit) can then determine whether dietary consumption of red meat and eggs has a statistically significant effect on serum ferritin concentrations (after adjusting for age and gender).

(Choice A) The analysis of variance (ANOVA) test compares the mean of a quantitative variable of ≥3 independent groups. For example, it could

- The quantitative dependent variable was serum ferritin concentration.
- The independent variables were consumption of red meat and eggs reported in ounces/week (2 quantitative variables).
- The effect of the main independent variables (eg, consumption of red meat and eggs) on the dependent variable (eg, serum ferritin concentration) is adjusted to account for the effects of other independent variables, which are called adjustment variables (eg, age, gender). Controlling for these variables reduces the potential confounding.

Calculation of a regression line (eg, line of best fit) can then determine whether dietary consumption of red meat and eggs has a statistically significant effect on serum ferritin concentrations (after adjusting for age and gender).

(Choice A) The analysis of variance (ANOVA) test compares the mean of a quantitative variable of ≥3 independent groups. For example, it could be used in a study comparing serum ferritin concentrations (ie, quantitative variable) in children (age 0-12), adolescents (age 13-17), adults (age 18-59), and seniors (age ≥60).

(Choice B) Meta-analysis is a quantitative statistical technique used to combine and analyze data from several studies to conduct an analysis with a greater statistical power than that of the individual studies.

(Choices C and E) The odds ratio (OR) and the relative risk (RR) measure the strength of association between 2 categorical variables: a categorical risk factor (eg, smoker or nonsmoker) and a categorical outcome/disease (eg, presence or absence of myocardial infarction). The OR is used in case-control studies whereas the RR is used in cohort studies.

Educational objective:
A regression analysis is a statistical technique used to describe the effect that 1 or more independent variables (eg, exposures, risk factors), which may be quantitative or qualitative, can have on 1 quantitative dependent variable (ie, outcome).

References

- Statistics for clinicians: an introduction to linear regression.
- Red meat and egg intake and serum ferritin concentrations in Colombian children: results of a population survey, ENSIN-2015.

Biostatistics Biostatistics & Epidemiology Statistical tests

A group of researchers wants to identify factors related to hospital-acquired bacteremia caused by methicillin-resistant *Staphylococcus aureus* (MRSA). A total of 45 patients were enrolled in the study after MRSA was isolated from a blood sample taken from them on the third or subsequent day after admission. In addition, 90 patients admitted to the hospital over the same period with a length of stay >2 days who did not have bacteremia were randomly selected. The frequency of factors such as insertion of a central line or urinary catheter and surgical site infection were then compared between the 2 groups. Which of the following is the most appropriate null hypothesis for this study?

A. Hazard ratio is equal to 1

B. Hazard ratio is not equal to 1

C. Odds ratio is equal to 1

D. Odds ratio is not equal to 1

E. Relative risk is equal to 1

F. Relative risk is not equal to 1

Submit

A group of researchers wants to identify factors related to hospital-acquired bacteremia caused by methicillin-resistant *Staphylococcus aureus* (MRSA). A total of 45 patients were enrolled in the study after MRSA was isolated from a blood sample taken from them on the third or subsequent day after admission. In addition, 90 patients admitted to the hospital over the same period with a length of stay >2 days who did not have bacteremia were randomly selected. The frequency of factors such as insertion of a central line or urinary catheter and surgical site infection were then compared between the 2 groups. Which of the following is the most appropriate null hypothesis for this study?

A. Hazard ratio is equal to 1 (4%)

B. Hazard ratio is not equal to 1 (2%)

✓ C. Odds ratio is equal to 1 (52%)

D. Odds ratio is not equal to 1 (9%)

E. Relative risk is equal to 1 (23%)

F. Relative risk is not equal to 1 (7%)

Omitted
Correct answer
C

52%
Answered correctly

02 secs
Time Spent

06/25/2020
Last Updated

Explanation

A statistical hypothesis is an initial assumption (that may or may not be true) regarding population parameters (eg, mean values in two different groups) or the relationship between variables within a population (eg, association between an exposure and a disease). The **null hypothesis (H_0)** is a statement of no difference or **no association**. The alternative hypothesis (H_a) is a statement of difference or association.

In this case, patients are **initially identified** based on their **disease status** (ie, cases: patients with methicillin-resistant *Staphylococcus aureus*

A statistical hypothesis is an initial assumption (that may or may not be true) regarding population parameters (eg, mean values in two different groups) or the relationship between variables within a population (eg, association between an exposure and a disease). The **null hypothesis (H_0)** is a statement of no difference or **no association**. The alternative hypothesis (H_a) is a statement of difference or association.

In this case, patients are **initially identified** based on their **disease status** (ie, cases: patients with methicillin-resistant *Staphylococcus aureus* [MRSA] bacteremia, controls: patients without MRSA bacteremia). The **frequency of factors** (ie, exposure to insertion of a central line or urinary catheter, surgical site infection) is **subsequently compared** between cases and controls. Therefore, this is a **case-control study**; the most appropriate measure of association between an exposure and a disease in case-control studies is the **odds ratio (OR)**, which represents the odds of disease among exposed patients relative to nonexposed patients.

The OR may be interpreted as follows:

- **OR = 1.0 (null value)** indicates that the odds of exposure among cases are the same as the odds of exposure among controls; therefore, **exposure is not associated with the disease**.
- OR ≠ 1.0 indicates that the odds of exposure are lower (ie, OR <1) or higher (ie, OR >1) among cases than among controls; therefore, exposure is associated with the disease.

Because H_0 is a statement of no association, the most appropriate H_0 for the study is that the OR = 1 **(Choice D)**.

(Choices A, B, E, and F) Measures of association based on risk/incidence (eg, relative risk, hazard ratio) are calculated in cohort studies and experimental designs, in which participants are initially selected based on exposure status (ie, exposed or nonexposed to a risk factor or treatment) and then are followed over time to assess development of disease. However, in case-control studies, *cases* (by definition) *already have the disease*. Therefore, case-control studies cannot calculate and compare the risk/incidence of disease between the case and control groups. Instead, case-control studies must compare the odds of exposure (to a risk modifier) between cases and controls.

Educational objective:
The odds ratio (OR) is a measure of association used in case-control studies. It quantifies the relationship between an exposure and a disease; its null value (ie, null hypothesis value) is always 1 (ie, OR = 1).

Biostatistics	Biostatistics & Epidemiology	Hypothesis testing
Subject	System	Topic

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Inflammatory biological markers are clinically useful in a number of different ways, one being the assessment of disease activity in conditions such as systemic lupus erythematosus (SLE). A new inflammatory marker is being investigated in patients with active SLE flares. When the blood level of the new marker (in mg/L) is plotted against the C-reactive protein (CRP) level (also in mg/L), the following plot is obtained:

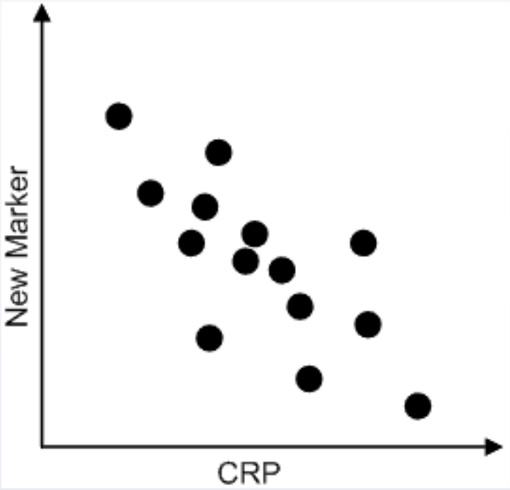
Based on the plot, the correlation coefficient between the 2 variables is closest to which of the following values?

A. +0.8
B. +0.2
C. 0
D. −0.2
E. −0.8

Submit

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Inflammatory biological markers are clinically useful in a number of different ways, one being the assessment of disease activity in conditions such as systemic lupus erythematosus (SLE). A new inflammatory marker is being investigated in patients with active SLE flares. When the blood level of the new marker (in mg/L) is plotted against the C-reactive protein (CRP) level (also in mg/L), the following plot is obtained:



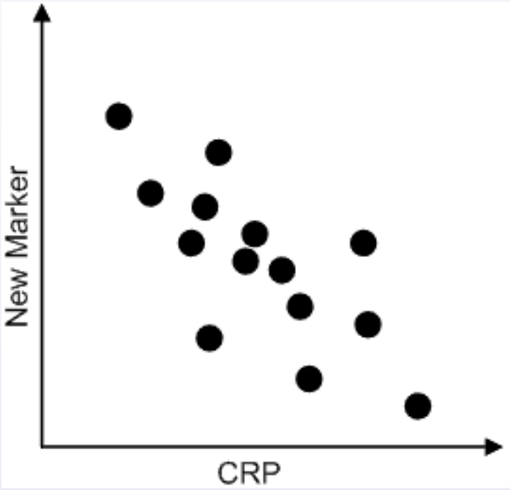
Based on the plot, the correlation coefficient between the 2 variables is closest to which of the following values?

A. +0.8
B. +0.2
C. 0
D. −0.2
E. −0.8

Submit

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

Inflammatory biological markers are clinically useful in a number of different ways, one being the assessment of disease activity in conditions such as systemic lupus erythematosus (SLE). A new inflammatory marker is being investigated in patients with active SLE flares. When the blood level of the new marker (in mg/L) is plotted against the C-reactive protein (CRP) level (also in mg/L), the following plot is obtained:

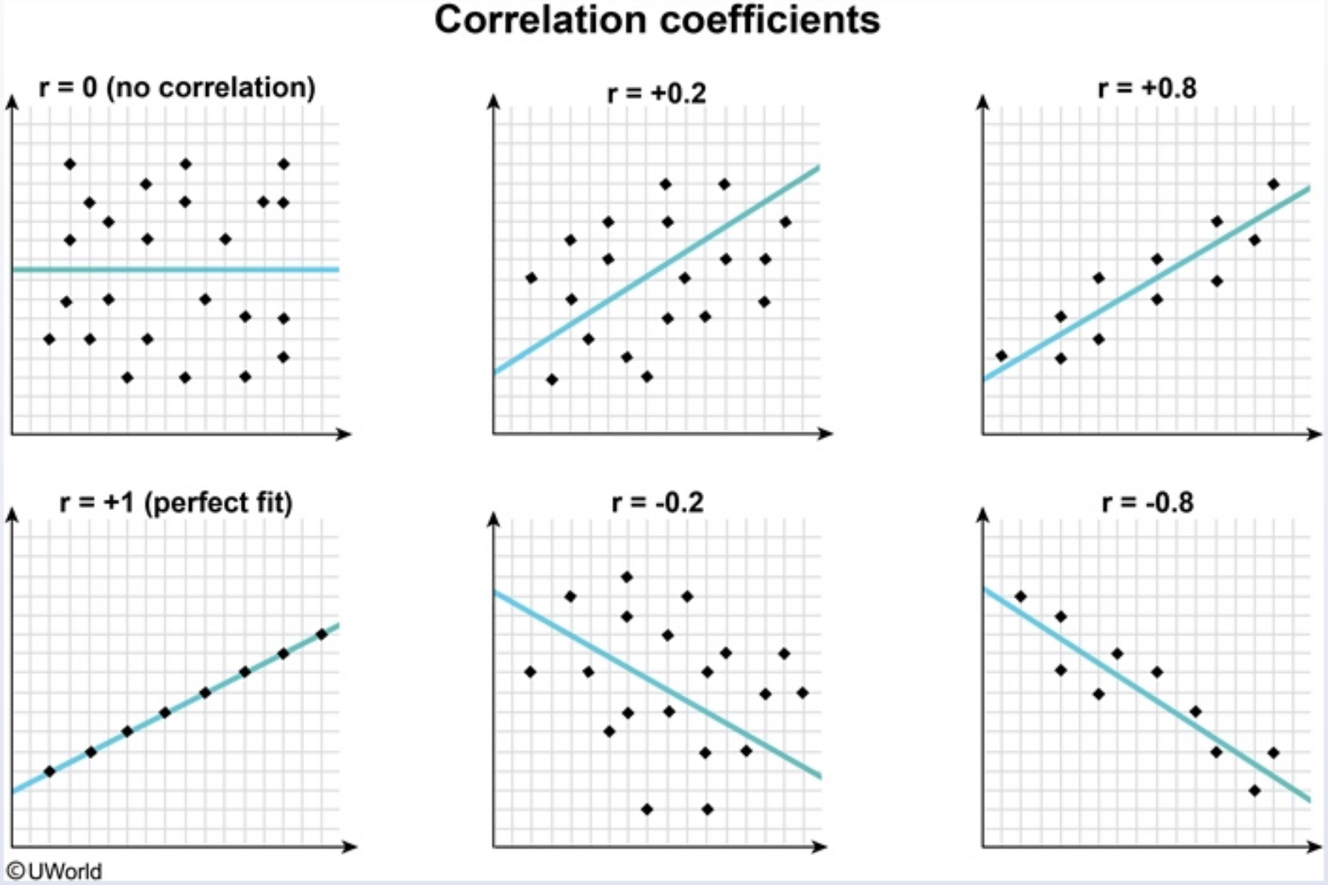


Based on the plot, the correlation coefficient between the 2 variables is closest to which of the following values?

A. +0.8 (4%)
B. +0.2 (5%)
C. 0 (6%)
D. −0.2 (10%)
E. −0.8 (73%)

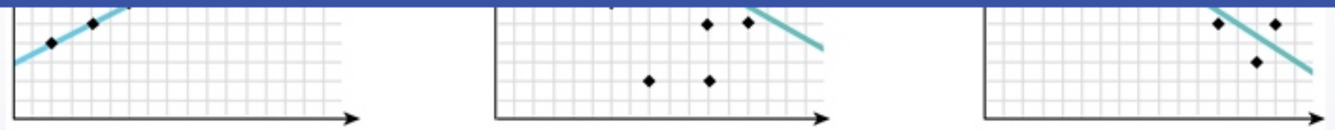
Omitted

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31



Scatter plots are useful for crude data analysis. If a **linear** association is present between 2 variables, a **correlation coefficient** (r) mathematically describes how well a "line of best fit" (blue line in Figure) would correspond to the data points plotted. The value of r ranges from **−1 to +1** and describes 2 important characteristics of an association: the strength and the polarity. The closer the r value is to its margins [−1, 1], the stronger the association.

An increase in C-reactive protein (CRP) level is associated with a decrease in the new marker's level; therefore, r <0. Although the scatter plot does not demonstrate a perfect linear arrangement, it does show a reasonably strong linear association as there is minimal variation along the line



©UWorld

Scatter plots are useful for crude data analysis. If a **linear** association is present between 2 variables, a **correlation coefficient** (r) mathematically describes how well a "line of best fit" (blue line in Figure) would correspond to the data points plotted. The value of r ranges from **−1 to +1** and describes 2 important characteristics of an association: the strength and the polarity. The closer the r value is to its margins [−1, 1], the stronger the association.

An increase in C-reactive protein (CRP) level is associated with a decrease in the new marker's level; therefore, r <0. Although the scatter plot does not demonstrate a perfect linear arrangement, it does show a reasonably strong linear association as there is minimal variation along the line of best fit (points are relatively close to the line). Therefore, the most appropriate answer among the options given is r = −0.8.

(Choices A and B) In this example, r cannot be >0 as an increase in CRP level is associated with a decrease in the new marker's level.

(Choice C) An r = 0 indicates that there is no association (ie, a random distribution).

(Choice D) An r = −0.2 would also indicate a negative association; however, there would be more variation along the line of best fit (as seen in the Figure).

The value of r is **not** the slope of the line of best fit. For instance, assuming a data set with a maximum positive association (ie, all data points forming a straight line with a positive slope), the line of best fit will always have r = +1, even if its actual slope is 0.2, 5.5, or any other positive value (as seen in the Figure).

Educational objective:
The correlation coefficient (r) ranges from −1 to +1 and describes the strength and polarity of a linear association.

Biostatistics	Biostatistics & Epidemiology	Correlation coefficient
Subject	System	Topic

Copyright © UWorld. All rights reserved.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

A 28-year-old woman comes to the office to establish care. She recently moved to New York to begin a job as a copy editor at a major newspaper. She has no known medical problems. The patient is in a monogamous relationship with her husband. She usually eats fast food for lunch, and she and her husband cook in the evenings. She goes to the gym about once every 2 weeks. The patient's temperature is 36.7 C (98 F), blood pressure is 118/64 mm Hg, pulse is 60/min, and respirations are 14/min. Her BMI is 24.6 kg/m^2. Physical examination is unremarkable. Laboratory results are within the normal range. Counseling this patient regarding diet and exercise habits would be an example of which of the following?

A. Case finding
B. Cognitive-behavioral therapy
C. Community-level intervention
D. Health promotion
E. Health risk assessment
F. Precontemplative stage intervention
G. Tertiary prevention

Submit

A 28-year-old woman comes to the office to establish care. She recently moved to New York to begin a job as a copy editor at a major newspaper. She has no known medical problems. The patient is in a monogamous relationship with her husband. She usually eats fast food for lunch, and she and her husband cook in the evenings. She goes to the gym about once every 2 weeks. The patient's temperature is 36.7 C (98 F), blood pressure is 118/64 mm Hg, pulse is 60/min, and respirations are 14/min. Her BMI is 24.6 kg/m^2. Physical examination is unremarkable. Laboratory results are within the normal range. Counseling this patient regarding diet and exercise habits would be an example of which of the following?

- A. Case finding (0%)
- B. Cognitive-behavioral therapy (0%)
- C. Community-level intervention (1%)
- ✓ D. Health promotion (81%)
- E. Health risk assessment (5%)
- F. Precontemplative stage intervention (5%)
- G. Tertiary prevention (5%)

Omitted
Correct answer
D

81%
Answered correctly

02 secs
Time Spent

04/19/2020
Last Updated

Explanation

Primary, secondary & tertiary prevention		
Level	**Definition**	**Example(s)**
Primary	Preventing a disease process from becoming established	• Health promotion (eg, regular exercise, no smoking, weight loss)
Secondary	Detecting a disease process before it causes symptoms	• Individual case finding (eg, cervical cancer screening) • Community screening (eg, blood pressure screening at state fair)
Tertiary	Treating a disease to prevent progression/complications	• Disability limitation (eg, blood sugar and blood pressure control in diabetes) • Rehabilitation (eg, physical therapy after stroke)

©UWorld

According to the World Health Organization, **health promotion** is "the process of enabling people to increase control over their health and its determinants, and thereby improve their health". Examples include improving dietary habits (limiting fat intake, consuming fruits and vegetables), exercising regularly, abstaining from smoking, and losing weight if needed. Health promotion typically falls under **primary** prevention, which is preventing a disease process from getting established.

Secondary prevention refers to interrupting the disease process before symptoms develop. An example is case finding, which is looking for disease that may be present but **asymptomatic** in patients receiving medical care (eg, age- and gender-appropriate screening) **(Choice A)**.

exercising regularly, abstaining from smoking, and losing weight if needed. Health promotion typically falls under **primary** prevention, which is preventing a disease process from getting established.

Secondary prevention refers to interrupting the disease process before symptoms develop. An example is case finding, which is looking for disease that may be present but **asymptomatic** in patients receiving medical care (eg, age- and gender-appropriate screening) **(Choice A)**. Community screening is an analogous intervention at the community level (eg, measuring hypertension at a state fair). **Tertiary** prevention corresponds to treating an established condition with the goal of minimizing its progression or complications **(Choice G)**.

(Choice B) Cognitive-behavioral therapy is a standardized psychotherapeutic modality focusing on identifying and addressing persistent maladaptive thought patterns. It is used as monotherapy or in combination with medication to treat a variety of mental illnesses (eg, anxiety, mood, personality, somatic symptom, eating disorders).

(Choice C) Interventions such as imposing taxes on cigarettes or soda or mandating smoke-free establishments can be implemented at the community level to improve the health of the public.

(Choice E) Health risk assessments rely on questionnaires that use demographic, medical, lifestyle, and family history information to calculate a patient's "risk age". A patient whose risk age is greater than chronological age has a higher risk of death than the average individual of that chronological age.

(Choice F) The "stages of change" model (originally developed to help patients quit smoking) assesses an individual's readiness to change a problem behavior. The precontemplative stage is when a patient does not believe there is a problem (eg, smoking) and is unwilling to contemplate a change (eg, quitting). The subsequent stages are contemplation, preparation, action, and maintenance.

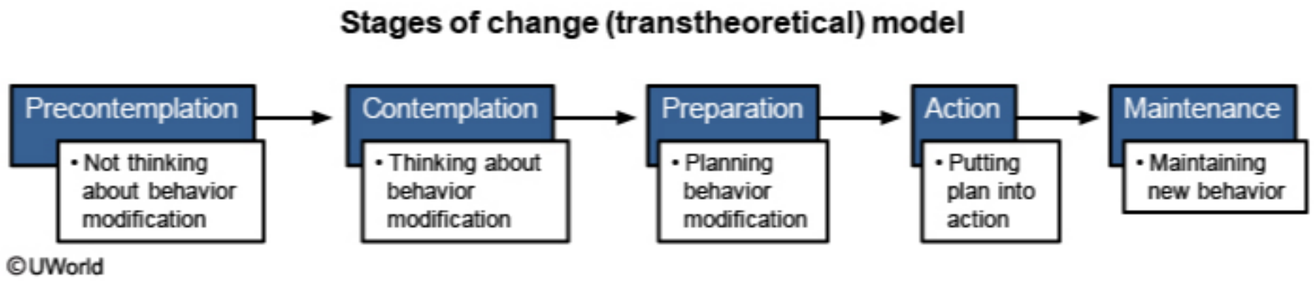
Educational objective:
Health promotion, according to the World Health Organization, is "the process of enabling people to increase control over their health and its determinants, and thereby improve their health."

Biostatistics	Biostatistics & Epidemiology	Prevention levels
Subject	System	Topic

Copyright © UWorld. All rights reserved.

exercising regularly, abstaining from smoking, and losing weight if needed. Health promotion typically falls under **primary** prevention, which is

Exhibit Display



Zoom In | Zoom Out | Reset | Add To New Card | Existing Card

Researchers conducted a randomized controlled trial to assess the effectiveness of a new drug to reduce severe cutaneous reactions in patients with rheumatoid arthritis who are treated with adalimumab. Of 150 adults with rheumatoid arthritis who are treated with adalimumab, 75 received the new drug in addition to adalimumab, and 75 received a placebo in addition to adalimumab. Results show that 6 patients in the new drug group developed severe cutaneous reactions, compared to 9 in the placebo group. Which of the following represents the relative risk reduction for severe cutaneous reactions among patients in the new drug group?

A. 0.08

B. 0.10

C. 0.12

D. 0.33

E. 0.67

Submit

Researchers conducted a randomized controlled trial to assess the effectiveness of a new drug to reduce severe cutaneous reactions in patients with rheumatoid arthritis who are treated with adalimumab. Of 150 adults with rheumatoid arthritis who are treated with adalimumab, 75 received the new drug in addition to adalimumab, and 75 received a placebo in addition to adalimumab. Results show that 6 patients in the new drug group developed severe cutaneous reactions, compared to 9 in the placebo group. Which of the following represents the relative risk reduction for severe cutaneous reactions among patients in the new drug group?

- A. 0.08 (9%)
- B. 0.10 (2%)
- C. 0.12 (4%)
- ✓ D. 0.33 (61%)
- E. 0.67 (22%)

Omitted
Correct answer
D

61%
Answered correctly

02 secs
Time Spent

05/13/2020
Last Updated

Explanation

Relative risk reduction (RRR) measures how much a given **treatment reduces the risk of an unfavorable outcome** relative to the control group. RRR may be defined as the difference in risk of the unfavorable outcome in the treatment group (eg, new drug) compared to the control group (eg, placebo):

$$RRR = (Risk_{control} - Risk_{treatment}) / Risk_{control}$$

In this example, Risk (ie, risk of severe cutaneous reactions in the placebo group) is **9/75 = 0.12**, and Risk (ie, risk of severe cutaneous

Relative risk reduction (RRR) measures how much a given **treatment reduces the risk of an unfavorable outcome** relative to the control group. RRR may be defined as the difference in risk of the unfavorable outcome in the treatment group (eg, new drug) compared to the control group (eg, placebo):

$$RRR = (Risk_{control} - Risk_{treatment}) / Risk_{control}$$

In this example, $Risk_{control}$ (ie, risk of severe cutaneous reactions in the placebo group) is **9/75 = 0.12**, and $Risk_{treatment}$ (ie, risk of severe cutaneous reactions in the new drug group) is **6/75 = 0.08**. Therefore, RRR for severe cutaneous reactions among patients in the new drug group is:

$$\mathbf{RRR = (0.12 - 0.08) / 0.12 = 0.33}$$

Alternately, RRR may be calculated using relative risk (RR). When using RR, the formula for RRR is:

$$RRR = 1 - RR$$

For this example, RR is the ratio of the risk of severe cutaneous reactions in the new drug group to the risk of severe cutaneous reactions in the placebo group, calculated as follows:

$$RR = (Risk_{treatment} / Risk_{control}) = (0.08/0.12) = 0.67$$

Therefore, RRR is:

$$RRR = 1 - 0.67 = 0.33$$

(Choice A) 0.08 represents the risk of severe cutaneous reactions in the new drug group (6/75 = 0.08, or 8%), not RRR.

(Choice B) 0.10 represents the risk of severe cutaneous reactions in the entire study (ie, both groups: [6 + 9] / [75 + 75] = 0.10, or 10%).

(Choice C) 0.12 represents the risk of severe cutaneous reactions in the placebo group (9/75 = 0.12, or 12%).

(Choice E) 0.67 represents the RR, the ratio of the risk of severe cutaneous reactions in the new drug group to the risk of severe cutaneous reactions in the placebo group (0.08/0.12 = 0.67, or 67%).

Educational objective:
Relative risk reduction (RRR) measures how much a treatment reduces the risk of an unfavorable outcome. Two methods to calculate it are:

For this example, RR is the ratio of the risk of severe cutaneous reactions in the new drug group to the risk of severe cutaneous reactions in the placebo group, calculated as follows:

$$RR = (Risk_{treatment} / Risk_{control}) = (0.08/0.12) = 0.67$$

Therefore, RRR is:

$$RRR = 1 - 0.67 = 0.33$$

(Choice A) 0.08 represents the risk of severe cutaneous reactions in the new drug group (6/75 = 0.08, or 8%), not RRR.

(Choice B) 0.10 represents the risk of severe cutaneous reactions in the entire study (ie, both groups: [6 + 9] / [75 + 75] = 0.10, or 10%).

(Choice C) 0.12 represents the risk of severe cutaneous reactions in the placebo group (9/75 = 0.12, or 12%).

(Choice E) 0.67 represents the RR, the ratio of the risk of severe cutaneous reactions in the new drug group to the risk of severe cutaneous reactions in the placebo group (0.08/0.12 = 0.67, or 67%).

Educational objective:
Relative risk reduction (RRR) measures how much a treatment reduces the risk of an unfavorable outcome. Two methods to calculate it are:

- $RRR = (Risk_{control} - Risk_{treatment}) / Risk_{control}$
- RRR = 1 − relative risk

References

- Measures of effect in epidemiological research.
- Measures of effect: relative risks, odds ratios, risk difference, and 'number needed to treat.'

Biostatistics	Biostatistics & Epidemiology	Measures of association
Subject	System	Topic

Copyright © UWorld. All rights reserved.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Researchers want to estimate the association between environmental lead exposure and cognitive deficits in children. Among all children who received preventive care at 5 local pediatric clinics over the last 5 years, they identify 20 children diagnosed with cognitive deficits and 40 without the diagnosis. The researchers then review the patients' medical records and calculate the distance between the patients' residence and known lead-contaminated geographic areas. Which of the following best describes the study design used by the researchers?

A. Case-control study

B. Case series study

C. Cross-sectional study

D. Prospective cohort study

E. Randomized control trial

Submit

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Researchers want to estimate the association between environmental lead exposure and cognitive deficits in children. Among all children who received preventive care at 5 local pediatric clinics over the last 5 years, they identify 20 children diagnosed with cognitive deficits and 40 without the diagnosis. The researchers then review the patients' medical records and calculate the distance between the patients' residence and known lead-contaminated geographic areas. Which of the following best describes the study design used by the researchers?

- ✓ A. Case-control study (68%)
- B. Case series study (10%)
- C. Cross-sectional study (16%)
- D. Prospective cohort study (4%)
- E. Randomized control trial (0%)

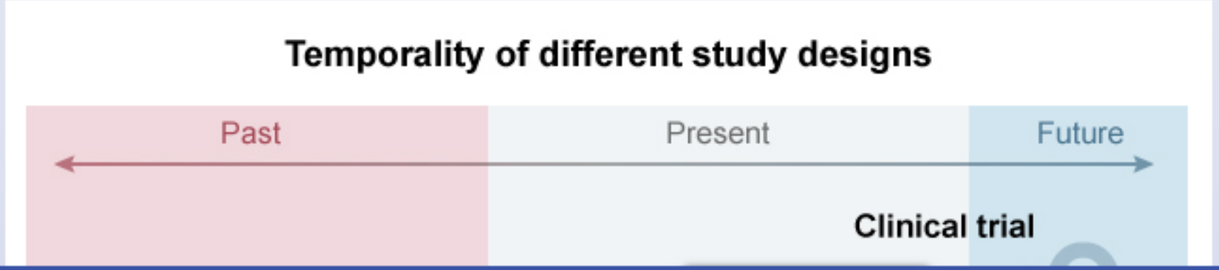
Omitted
Correct answer
A

68%
Answered correctly

01 sec
Time Spent

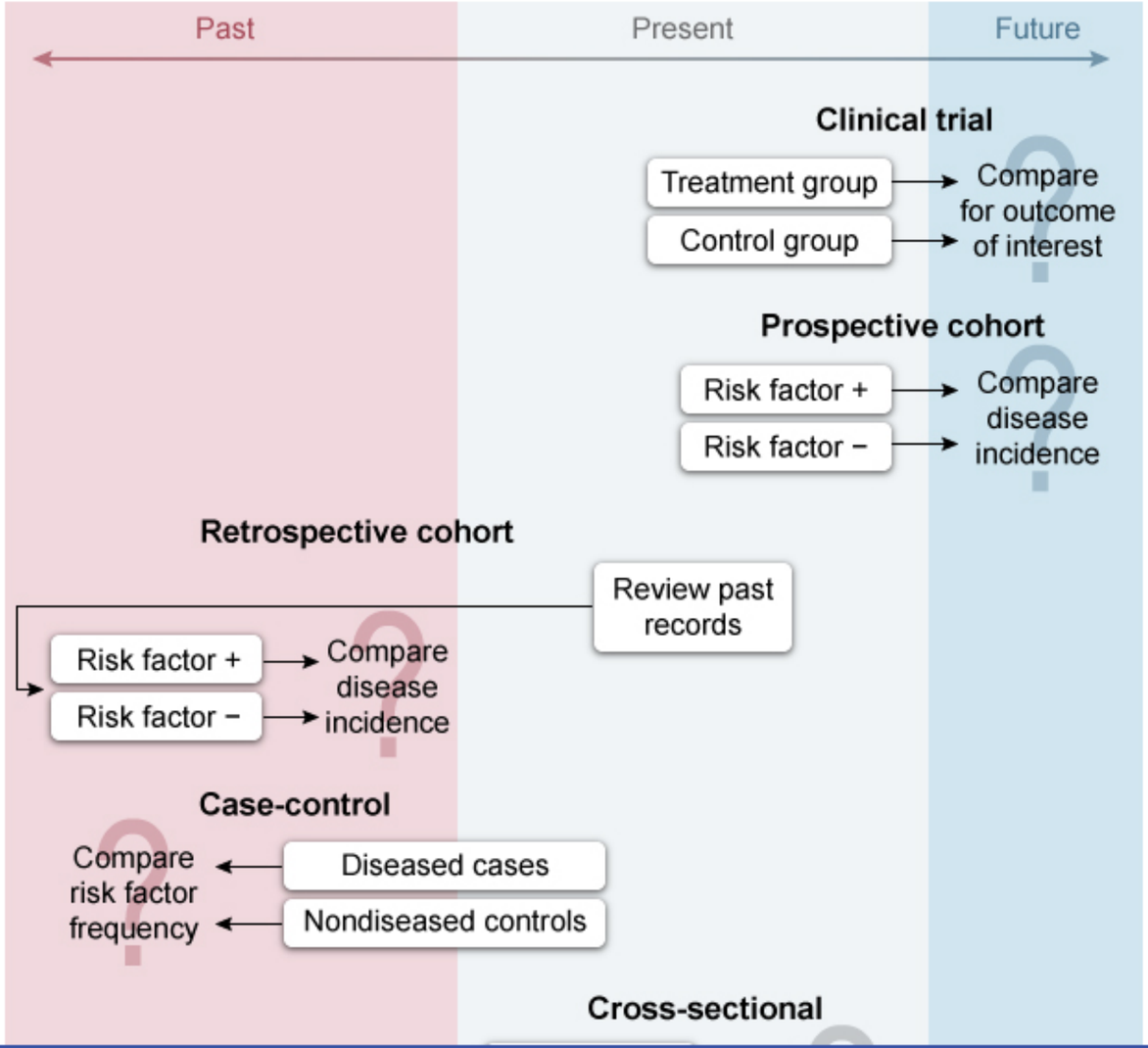
05/09/2020
Last Updated

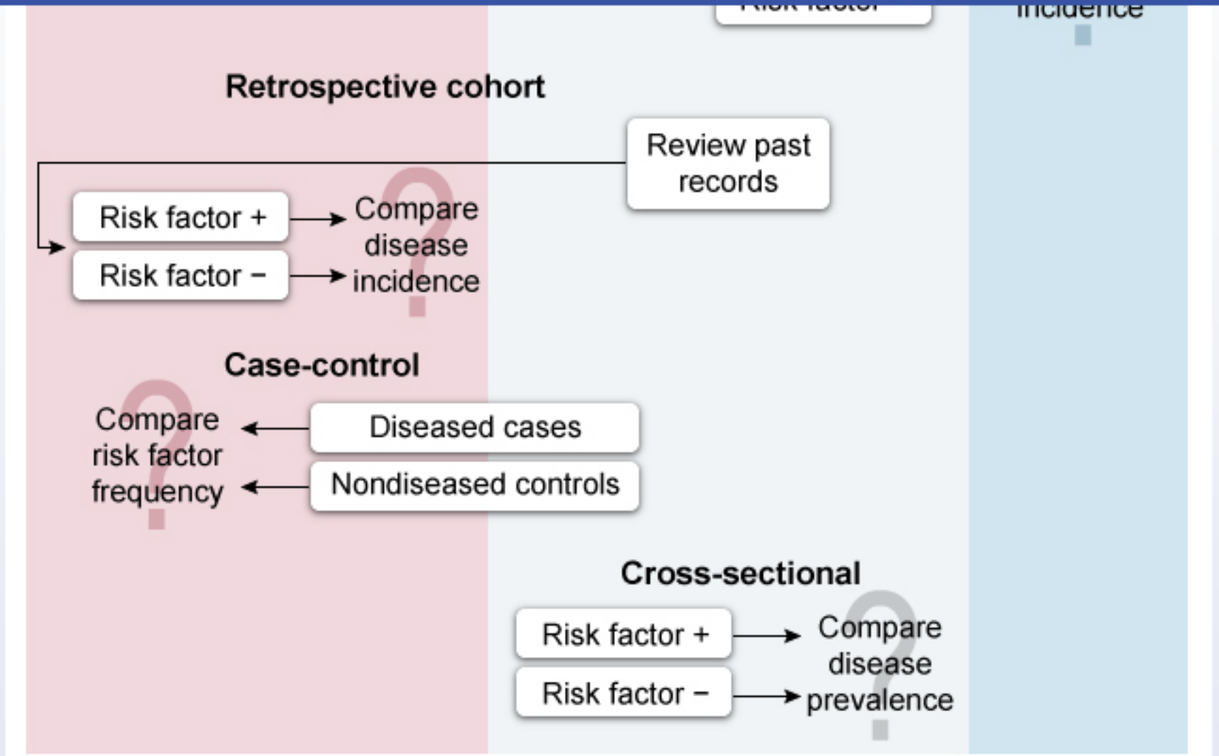
Explanation



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Temporality of different study designs





Observational studies (eg, case series, case-control studies, cohort studies, cross-sectional studies) differ from experimental studies (eg, clinical trials) in that the researchers passively observe the study participants rather than monitor the result of a specific intervention. Observational studies can be subdivided as follows:

- Descriptive studies collect data to characterize a specific health problem (eg, disease prevalence, incidence rates).
- Analytical studies test hypotheses to evaluate associations between risk factors and disease.

A **case-control** study is an analytical observational design in which potential participants are initially identified as "**cases**" or "**controls**" according to an **outcome status** (ie, disease of interest). In this example, children diagnosed with cognitive deficits are cases and children without cognitive deficits are controls. Once identified, cases and controls are then **assessed retrospectively** for exposure to **risk factors** of interest. In this case,

©UWorld

Observational studies (eg, case series, case-control studies, cohort studies, cross-sectional studies) differ from experimental studies (eg, clinical trials) in that the researchers passively observe the study participants rather than monitor the result of a specific intervention. Observational studies can be subdivided as follows:

- Descriptive studies collect data to characterize a specific health problem (eg, disease prevalence, incidence rates).
- Analytical studies test hypotheses to evaluate associations between risk factors and disease.

A **case-control** study is an analytical observational design in which potential participants are initially identified as "**cases**" or "**controls**" according to an **outcome status** (ie, disease of interest). In this example, children diagnosed with cognitive deficits are cases and children without cognitive deficits are controls. Once identified, cases and controls are then **assessed retrospectively** for exposure to **risk factors** of interest. In this case, the risk factor of interest is environmental lead exposure, as measured by the distance between the patients' residence and known lead-contaminated areas. Finally, the prevalence of exposure to the risk factors (ie, frequency of exposure) is compared between cases and controls to estimate the **association** between the risk factors and the outcomes.

(Choice B) A case series study is a descriptive study that tracks patients with a known condition (eg, exposure, risk factor, disease) to document the natural history or response to treatment. Unlike a case-control study, a case series is purely descriptive and cannot establish associations.

(Choice C) Cross-sectional studies may evaluate associations between risk factors and outcomes of interest. However, unlike in case-control studies (participants selected based on the outcome status), the participants in a cross-sectional study are selected based on the inclusion and exclusion criteria set for the study.

(Choice D) In cohort studies, 2 groups of individuals (ie, cohorts) are initially identified based on their exposure status to a specific risk factor (eg, environmental lead exposure) rather than based on their outcome status (eg, cognitive deficits), as seen in this example. The cohorts are then followed over time for development of the outcome.

(Choice E) A randomized clinical trial would follow individuals who have been randomized to either a treatment arm or a control arm and compare the effect of the intervention.

Educational objective:
A case-control study is an observational study design; it begins with selecting individuals who have the outcome ("cases") and individuals who do

- Analytical studies test hypotheses to evaluate associations between risk factors and disease.

A **case-control** study is an analytical observational design in which potential participants are initially identified as "**cases**" or "**controls**" according to an **outcome status** (ie, disease of interest). In this example, children diagnosed with cognitive deficits are cases and children without cognitive deficits are controls. Once identified, cases and controls are then **assessed retrospectively** for exposure to **risk factors** of interest. In this case, the risk factor of interest is environmental lead exposure, as measured by the distance between the patients' residence and known lead-contaminated areas. Finally, the prevalence of exposure to the risk factors (ie, frequency of exposure) is compared between cases and controls to estimate the **association** between the risk factors and the outcomes.

(Choice B) A case series study is a descriptive study that tracks patients with a known condition (eg, exposure, risk factor, disease) to document the natural history or response to treatment. Unlike a case-control study, a case series is purely descriptive and cannot establish associations.

(Choice C) Cross-sectional studies may evaluate associations between risk factors and outcomes of interest. However, unlike in case-control studies (participants selected based on the outcome status), the participants in a cross-sectional study are selected based on the inclusion and exclusion criteria set for the study.

(Choice D) In cohort studies, 2 groups of individuals (ie, cohorts) are initially identified based on their exposure status to a specific risk factor (eg, environmental lead exposure) rather than based on their outcome status (eg, cognitive deficits), as seen in this example. The cohorts are then followed over time for development of the outcome.

(Choice E) A randomized clinical trial would follow individuals who have been randomized to either a treatment arm or a control arm and compare the effect of the intervention.

Educational objective:
A case-control study is an observational study design; it begins with selecting individuals who have the outcome ("cases") and individuals who do not have the outcome ("controls") and then retrospectively comparing their history of exposure to risk factors.

Biostatistics	Biostatistics & Epidemiology	Study designs
Subject	System	Topic

Copyright © UWorld. All rights reserved.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Carbamoyl phosphate synthetase I deficiency is an inherited disorder characterized by accumulation of ammonia in the blood. The most severe form occurs in the first 24-72 hours following birth, after feeding begins and milk proteins start to be broken down in the liver. If left untreated, affected neonates often die due to severe metabolic derangements; survivors often develop permanent neurologic injury. The estimated incidence of carbamoyl phosphate synthetase I deficiency is about 1 in 800,000 newborns. If a decision is made to test all newborns for this disease, then this initial test should be designed to have a high:

A. Cutoff value

B. Number of true negatives

C. Positive predictive value

D. Sensitivity

E. Specificity

Submit

Carbamoyl phosphate synthetase I deficiency is an inherited disorder characterized by accumulation of ammonia in the blood. The most severe form occurs in the first 24-72 hours following birth, after feeding begins and milk proteins start to be broken down in the liver. If left untreated, affected neonates often die due to severe metabolic derangements; survivors often develop permanent neurologic injury. The estimated incidence of carbamoyl phosphate synthetase I deficiency is about 1 in 800,000 newborns. If a decision is made to test all newborns for this disease, then this initial test should be designed to have a high:

A. Cutoff value (0%)

B. Number of true negatives (2%)

C. Positive predictive value (7%)

✓ D. Sensitivity (77%)

E. Specificity (12%)

Omitted
Correct answer
D

77%
Answered correctly

01 sec
Time Spent

04/30/2020
Last Updated

Explanation

Carbamoyl phosphate synthetase I deficiency is a rare disorder (incidence 1 in 800,000 newborns) with serious and possible irreversible consequences if not detected and treated early. Therefore, it is important to identify all individuals who potentially have the disease using a **screening test** that can be performed on a large number of newborns.

Although an ideal screening test would be highly sensitive and specific, developing a test that has both of these properties is often not possible, and so a compromise must be made. The sensitivity of a test refers to its ability to correctly identify those with the disease: it is the probability of

Explanation

Carbamoyl phosphate synthetase I deficiency is a rare disorder (incidence 1 in 800,000 newborns) with serious and possible irreversible consequences if not detected and treated early. Therefore, it is important to identify all individuals who potentially have the disease using a **screening test** that can be performed on a large number of newborns.

Although an ideal screening test would be highly sensitive and specific, developing a test that has both of these properties is often not possible, and so a compromise must be made. The sensitivity of a test refers to its ability to correctly identify those with the disease: it is the probability of the test returning a positive result in a person with the disease. A **highly sensitive test** will ensure that most patients with the disease will have a positive test result (leading to few false negative results); therefore, **fewer cases of disease are missed**. Given a test with high sensitivity, a negative result would help to rule out a diagnosis (SnNout). This is important during screening for **life-threatening diseases**, even if obtaining a high sensitivity causes an increased numbers of false positives (ie, reduced specificity).

(Choice A) Setting a high cutoff value typically (but not always) produces higher specificity and lower sensitivity.

(Choices B and E) Specificity represents the ability of a test to correctly identify those without the disease. A very specific test has a low rate of false positives, so most healthy patients will have a negative test result (true negative). Given a test with high specificity, a positive result would help to rule in a diagnosis (SpPin). Confirmatory tests with high specificity are often used on patients who test positive on a screening test to ensure that a patient actually has the disease.

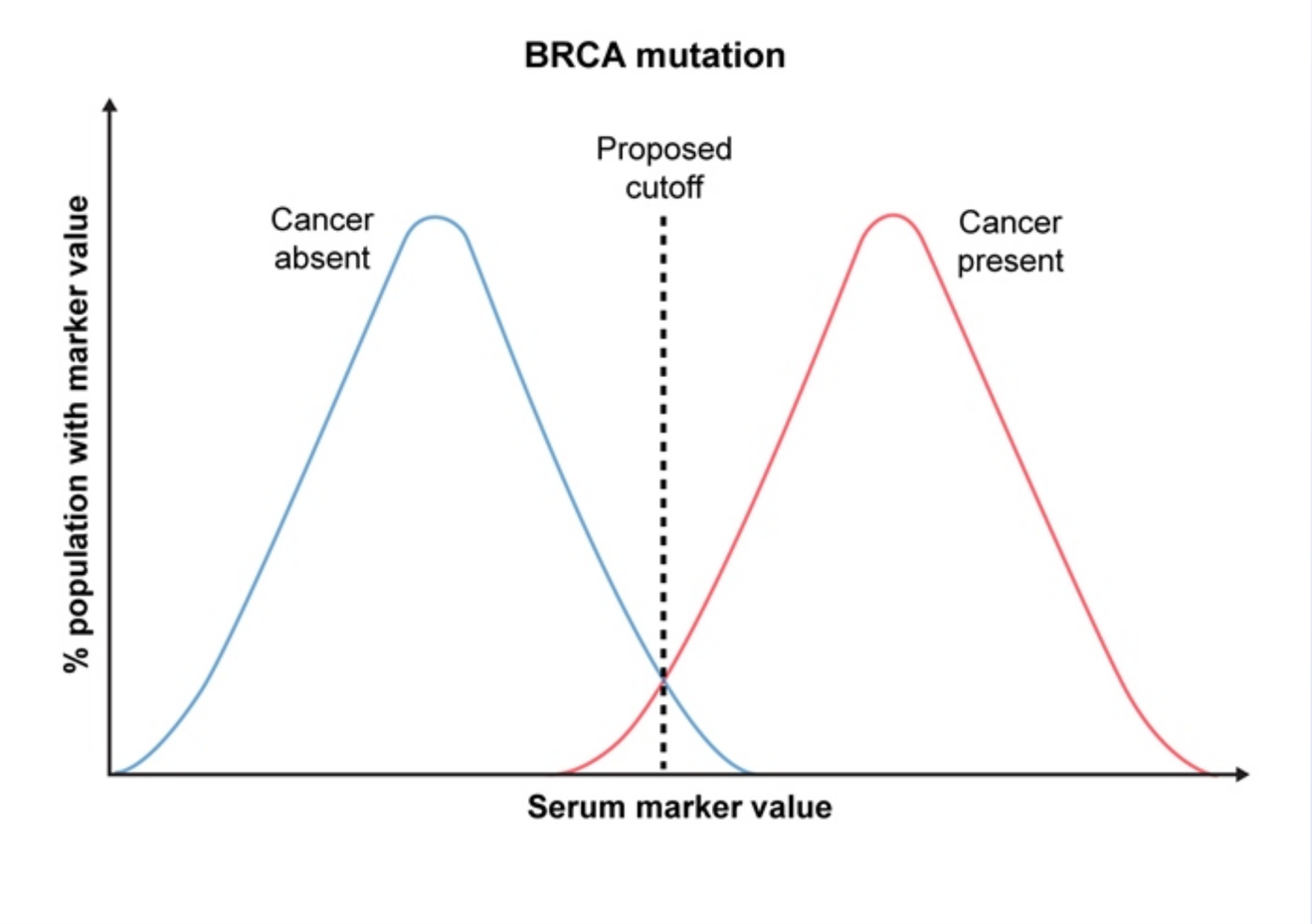
(Choice C) Positive predictive value refers to the probability that a disease is present given a positive test result. Positive predictive value depends on disease prevalence. When screening patients for a rare disorder, positive predictive value will often be low due to a high number of false positives.

Educational objective:
The sensitivity of a test refers to its ability to correctly identify those with the disease. A highly sensitive test should always be considered over a highly specific test when screening for life-threatening diseases, where identification of every person with the disease is important.

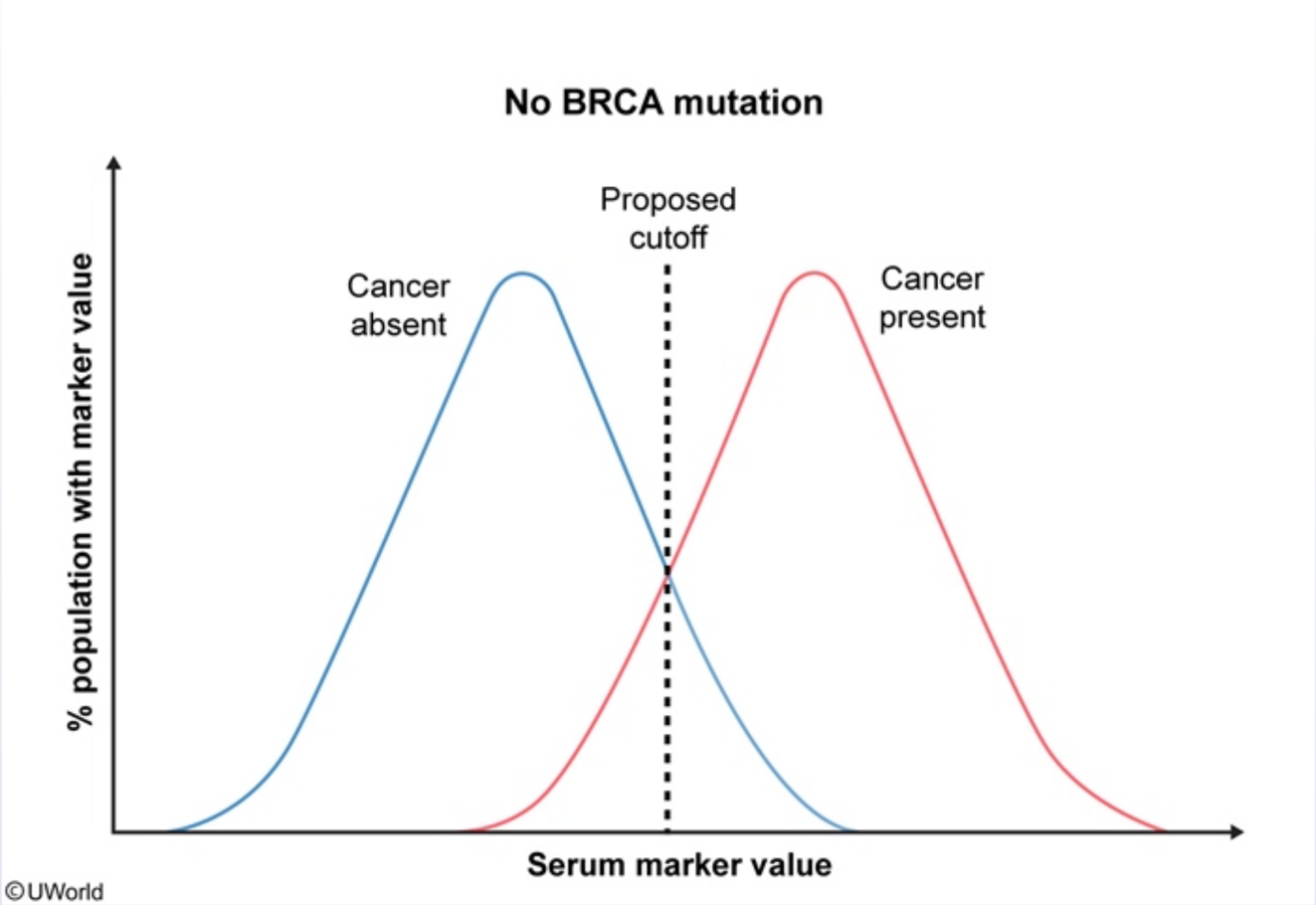
Biostatistics	Biostatistics & Epidemiology	Sensitivity and specificity
Subject	System	Topic

Copyright © UWorld. All rights reserved.

A new tumor marker is being investigated for its usefulness in diagnosing ovarian cancer. A sample of 400 women is stratified into 2 groups based on the presence or absence of *BRCA* mutations. Serum levels of the new marker are obtained in the 2 groups; in addition, both groups of women undergo conventional screening to determine their disease status. The curves on the top represent the distribution of the new serum marker in women with *BRCA* mutations, and the curves on the bottom represent the distribution of the new serum marker in women without the *BRCA* mutations.



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

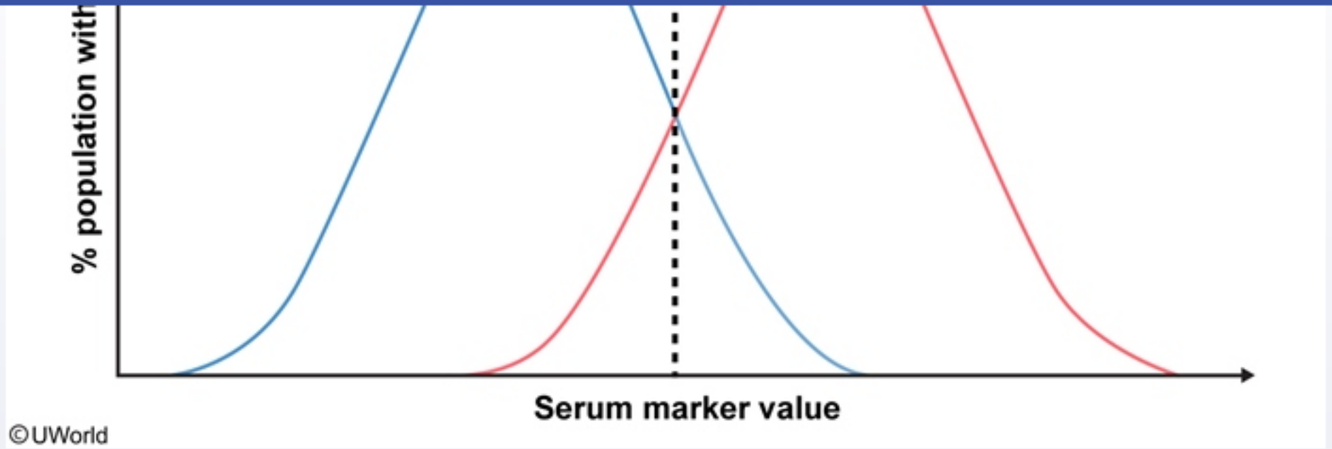


The proposed serum marker cutoff value for both populations is shown. Use of the new serum marker in women with *BRCA* mutations, compared with those without *BRCA* mutations, is associated with which of the following?

A. Higher sensitivity and higher specificity

B. Higher sensitivity and lower specificity

C. Higher sensitivity and same specificity

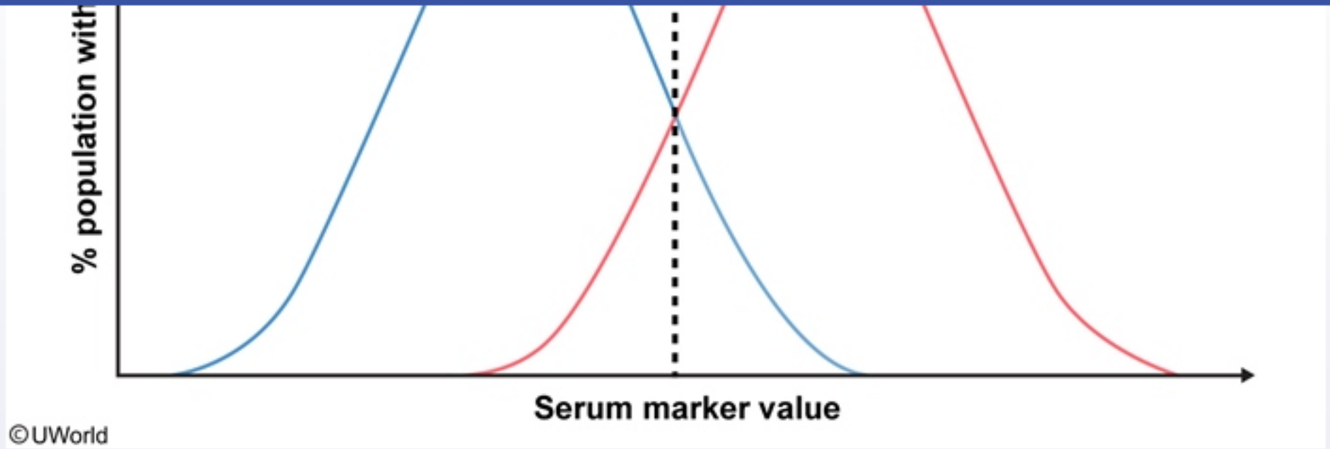


The proposed serum marker cutoff value for both populations is shown. Use of the new serum marker in women with *BRCA* mutations, compared with those without *BRCA* mutations, is associated with which of the following?

- A. Higher sensitivity and higher specificity
- B. Higher sensitivity and lower specificity
- C. Higher sensitivity and same specificity
- D. Lower sensitivity and higher specificity
- E. Lower sensitivity and lower specificity
- F. Lower sensitivity and unchanged specificity
- G. Unchanged sensitivity and unchanged specificity

Submit

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31



The proposed serum marker cutoff value for both populations is shown. Use of the new serum marker in women with *BRCA* mutations, compared with those without *BRCA* mutations, is associated with which of the following?

- ✓ A. Higher sensitivity and higher specificity (64%)
- B. Higher sensitivity and lower specificity (8%)
- C. Higher sensitivity and same specificity (1%)
- D. Lower sensitivity and higher specificity (7%)
- E. Lower sensitivity and lower specificity (11%)
- F. Lower sensitivity and unchanged specificity (0%)
- G. Unchanged sensitivity and unchanged specificity (5%)

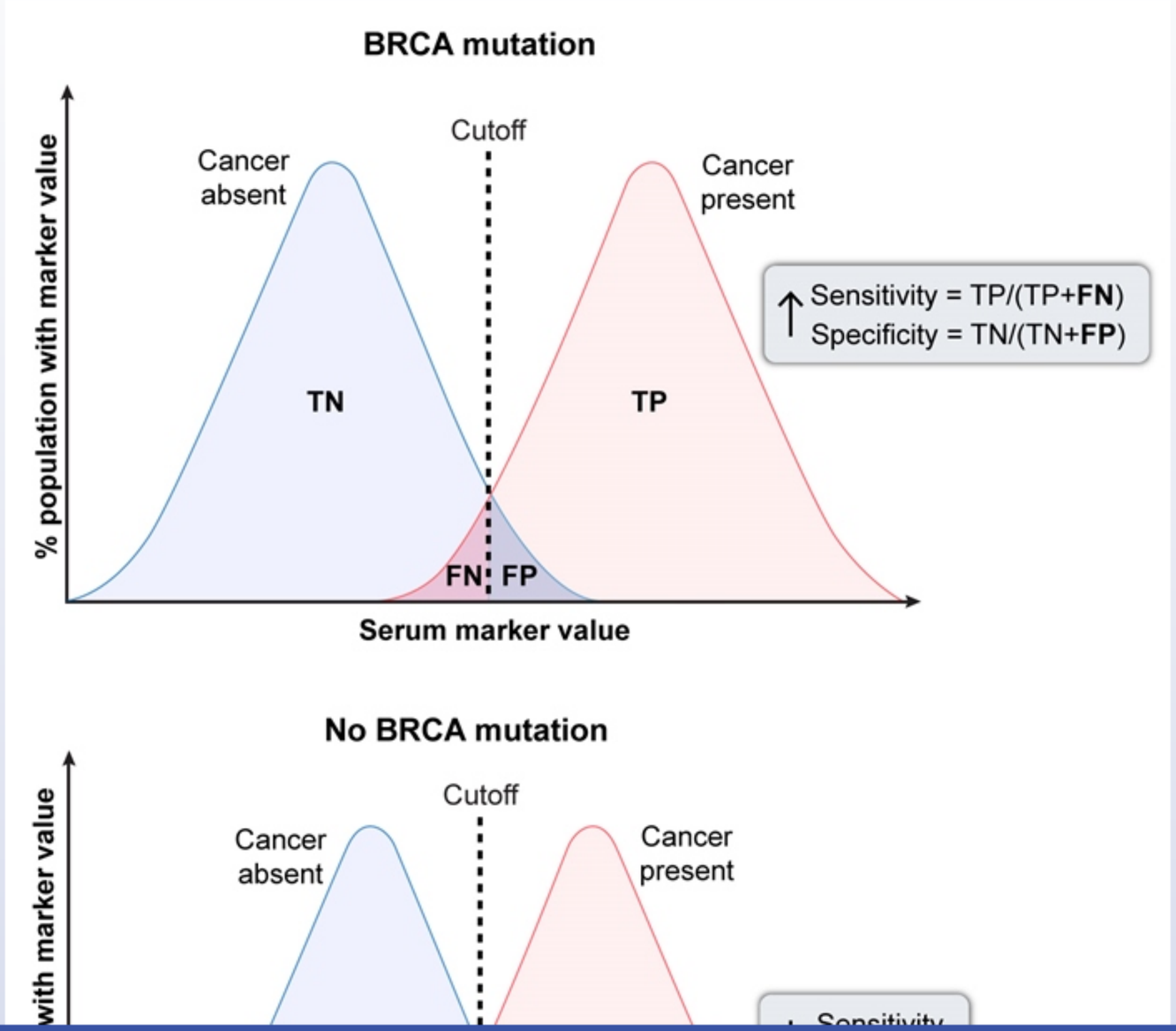
Omitted
Correct answer

64%
Answered correctly

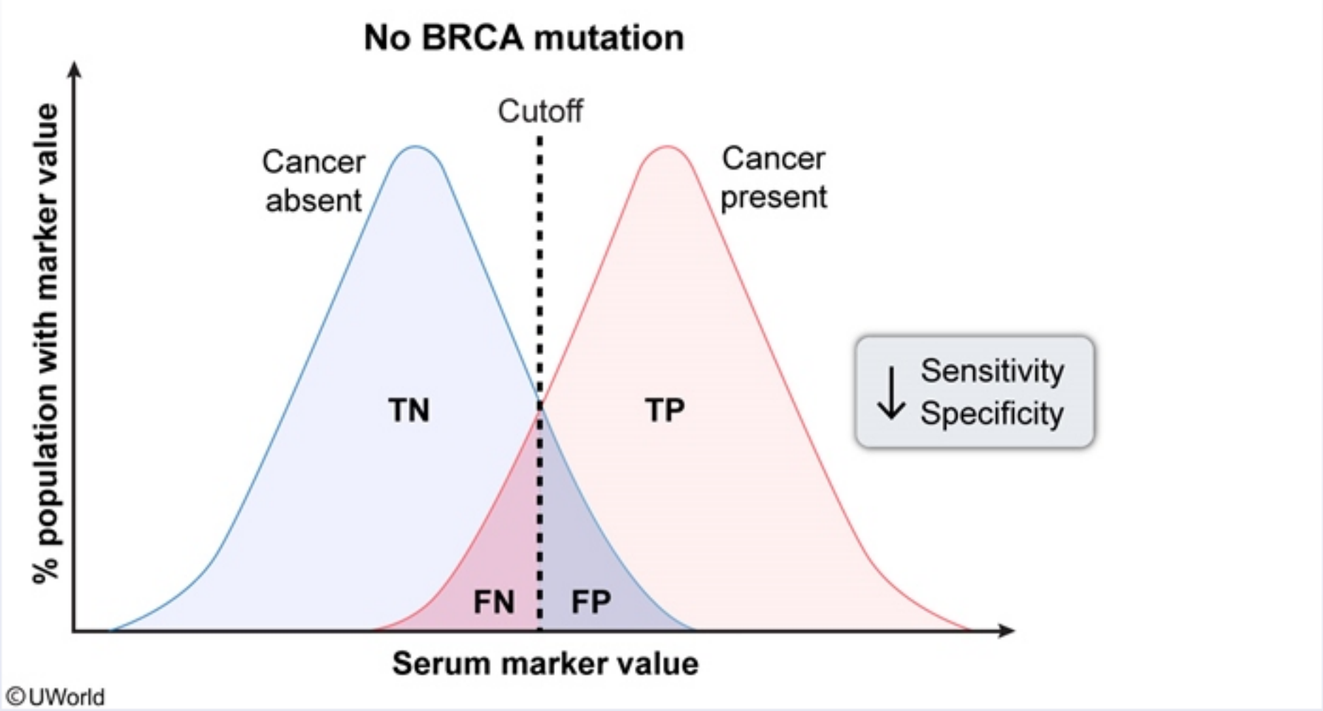
09 secs
Time Spent

05/08/2020
Last Updated

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31



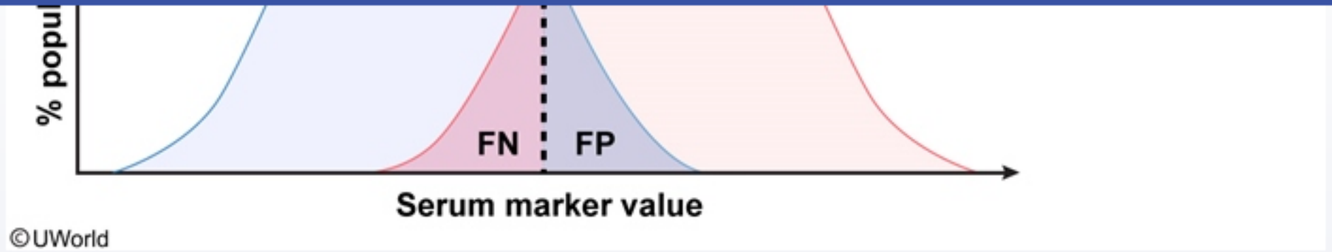
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31



The cutoff value of a quantitative diagnostic test determines whether a given result is interpreted as positive or negative. When there is **overlap between** the serum values of the **healthy and diseased populations**, a cutoff value that correctly categorizes all individuals in both populations cannot be chosen. This **limits the sensitivity and specificity** of the test due to the presence of false positive (FP) and/or false negative (FN) individuals.

Sensitivity represents the ability of a test to correctly identify those with a given disease. It is calculated as the number of patients correctly testing positive (TP) divided by the total number of patients with the disease (TP / [TP + FN]). Specificity represents the ability of a test to correctly identify those without a given disease. It is calculated as the number of patients correctly testing negative (TN) divided by the total number of patients without the disease (TN / [TN + FP]).

In this case, values of the serum marker in women with *BRCA* mutations show **decreased overlap** between the healthy and diseased curves of women with *BRCA* mutations (curves on the top) compared with women without *BRCA* mutations (curves on the bottom). The corresponding



The cutoff value of a quantitative diagnostic test determines whether a given result is interpreted as positive or negative. When there is **overlap between** the serum values of the **healthy and diseased populations**, a cutoff value that correctly categorizes all individuals in both populations cannot be chosen. This **limits the sensitivity and specificity** of the test due to the presence of false positive (FP) and/or false negative (FN) individuals.

Sensitivity represents the ability of a test to correctly identify those with a given disease. It is calculated as the number of patients correctly testing positive (TP) divided by the total number of patients with the disease (TP / [TP + FN]). Specificity represents the ability of a test to correctly identify those without a given disease. It is calculated as the number of patients correctly testing negative (TN) divided by the total number of patients without the disease (TN / [TN + FP]).

In this case, values of the serum marker in women with *BRCA* mutations show **decreased overlap** between the healthy and diseased curves of women with *BRCA* mutations (curves on the top) compared with women without *BRCA* mutations (curves on the bottom). The corresponding decrease in the number of FPs and FNs means the new serum marker has **higher sensitivity and specificity** (ie, better performance) in women with *BRCA* mutations.

Educational objective:
The degree of overlap between the healthy and the diseased population curves limits the maximum combined sensitivity and specificity of a quantitative diagnostic test. The degree to which sensitivity or specificity is affected depends on the chosen cutoff value.

Biostatistics	Biostatistics & Epidemiology	Sensitivity and specificity
Subject	System	Topic

Copyright © UWorld. All rights reserved.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

The aim of a prospective cohort study conducted in a group of adults with diagnosed type 2 diabetes mellitus (T2DM) is to investigate the association between the presence of morbid obesity (BMI >40 kg/m^2) and the risk of developing diabetic nephropathy. One of the groups in the study consists of adults with T2DM, a BMI >40 kg/m^2, and no diabetic nephropathy. Which of the following is the most appropriate comparison group for this prospective cohort study?

A. Adults with T2DM who have morbid obesity (BMI >40 kg/m^2) and diabetic nephropathy

B. Adults with T2DM who have normal weight (BMI 18.5 to <25 kg/m^2) and diabetic nephropathy

C. Adults with T2DM who have normal weight (BMI 18.5 to <25 kg/m^2) and no diabetic nephropathy

D. Adults without T2DM who have morbid obesity (BMI >40 kg/m^2) and no diabetic nephropathy

E. Adults without T2DM who have normal weight (BMI 18.5 to <25 kg/m^2) and diabetic nephropathy

Submit

The aim of a prospective cohort study conducted in a group of adults with diagnosed type 2 diabetes mellitus (T2DM) is to investigate the association between the presence of morbid obesity (BMI >40 kg/m^2) and the risk of developing diabetic nephropathy. One of the groups in the study consists of adults with T2DM, a BMI >40 kg/m^2, and no diabetic nephropathy. Which of the following is the most appropriate comparison group for this prospective cohort study?

A. Adults with T2DM who have morbid obesity (BMI >40 kg/m^2) and diabetic nephropathy (21%)

B. Adults with T2DM who have normal weight (BMI 18.5 to <25 kg/m^2) and diabetic nephropathy (6%)

✓ C. Adults with T2DM who have normal weight (BMI 18.5 to <25 kg/m^2) and no diabetic nephropathy (66%)

D. Adults without T2DM who have morbid obesity (BMI >40 kg/m^2) and no diabetic nephropathy (4%)

E. Adults without T2DM who have normal weight (BMI 18.5 to <25 kg/m^2) and diabetic nephropathy (1%)

Omitted
Correct answer
C

66%
Answered correctly

02 secs
Time Spent

05/15/2020
Last Updated

Explanation

A **cohort study** is an observational design in which potential participants in the **population of interest** are initially **identified as exposed or nonexposed** according to the independent variable (ie, exposure status to a risk factor). Once participants are categorized based on their exposure status, the occurrence (ie, incidence) of the dependent variable (eg, **outcome of interest**) over a specific period is determined in each group. Finally, the occurrence of the outcome is compared between exposed and nonexposed groups to estimate the association between the risk factors and the outcomes. If there is a statistically significant difference in outcome occurrence between the 2 groups, it is likely that the risk factor in question is associated with disease development.

A **cohort study** is an observational design in which potential participants in the **population of interest** are initially **identified as exposed or nonexposed** according to the independent variable (ie, exposure status to a risk factor). Once participants are categorized based on their exposure status, the occurrence (ie, incidence) of the dependent variable (eg, **outcome of interest**) over a specific period is determined in each group. Finally, the occurrence of the outcome is compared between exposed and nonexposed groups to estimate the association between the risk factors and the outcomes. If there is a statistically significant difference in outcome occurrence between the 2 groups, it is likely that the risk factor in question is associated with disease development.

In this example:

- The **population** of interest is patients with **type 2 diabetes mellitus** (T2DM). Therefore, both the exposed and nonexposed groups must consist of patients with T2DM **(Choices D and E)**.
- The **risk factor** of interest is **morbid obesity** (BMI >40 kg/m^2). Therefore, the exposed group is adults with morbid obesity, and the nonexposed group is adults with normal weight (BMI 18.5 to <25 kg/m^2) **(Choice A)**.
- The **outcome** of interest is **diabetic nephropathy**. Exposed and nonexposed subjects must be selected with **no history** of the outcome because the occurrence of the outcome is what is measured during the study period **(Choice B)**.

Therefore, the exposed group is adults with T2DM who have morbid obesity and no diabetic nephropathy; the **nonexposed group** is adults **with T2DM** who have **normal weight** (BMI 18.5 to <25 kg/m^2) and **no diabetic nephropathy**. Comparing the frequency of the outcome (eg, diabetic nephropathy) between exposed and nonexposed groups is what determines whether exposure to the risk factor (eg, morbid obesity) is associated with the risk of developing the outcome (eg, diabetic nephropathy).

Educational objective:
In a cohort study, subjects should be initially selected from the population of interest (eg, T2DM) based on their exposure status to a risk factor (eg, presence or absence of morbid obesity). Subjects are then monitored during the study period for development of the outcome of interest (eg, diabetic nephropathy).

References

- Methodology series module 1: cohort studies.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

A research group conducted a study to compare the levels of creatine kinase-MB (CK-MB) between type 2 diabetes mellitus (T2DM) patients given statin therapy. Participants were divided into 4 groups based on treatment. Groups I, II, and III consisted of T2DM patients who had been given statin therapy (atorvastatin, simvastatin, rosuvastatin, respectively) for at least 6 months. Group IV consisted of T2DM patients who had not been given statin therapy. Which of the following statistical tests is most appropriate to compare the CK-MB levels between Groups I, II, III, and IV in this study?

A. Analysis of variance

B. Chi-square test

C. Independent *t*-test

D. Paired *t*-test

E. Correlation analysis

Submit

A research group conducted a study to compare the levels of creatine kinase-MB (CK-MB) between type 2 diabetes mellitus (T2DM) patients given statin therapy. Participants were divided into 4 groups based on treatment. Groups I, II, and III consisted of T2DM patients who had been given statin therapy (atorvastatin, simvastatin, rosuvastatin, respectively) for at least 6 months. Group IV consisted of T2DM patients who had not been given statin therapy. Which of the following statistical tests is most appropriate to compare the CK-MB levels between Groups I, II, III, and IV in this study?

- ✓ A. Analysis of variance (77%)
- B. Chi-square test (11%)
- C. Independent *t*-test (2%)
- D. Paired *t*-test (3%)
- E. Correlation analysis (4%)

Omitted
Correct answer
A

77%
Answered correctly

03 secs
Time Spent

05/08/2020
Last Updated

Explanation

		Dependent variable	
		Qualitative (categorical)	*Quantitative*
Independent variable	*Qualitative (categorical)*	Chi-square, logistic regression*	*t*-test, ANOVA, linear regression

		Dependent variable	
		Qualitative (categorical)	*Quantitative*
Independent variable	*Qualitative (categorical)*	Chi-square, logistic regression*	*t*-test, ANOVA, linear regression
	Quantitative	Logistic regression*	Correlation, linear regression

* Dependent variable must be dichotomous.

ANOVA = analysis of variance.

Variables are broadly classified as qualitative (ie, categorical) or quantitative (eg, continuous) based on their scale of measurement. **Qualitative variables** (eg, type of treatment, blood type) represent categories or groups, whereas **quantitative variables** (eg, temperature, glucose levels) represent numerical values, with quantitative variables (eg, temperature) sometimes transformed into qualitative variables (eg, "no fever" for <38 C [100 F]; "fever" for ≥38 C [100 F]). The scale of measurement of the dependent (eg, outcome) and independent (eg, exposures, risk factors) variables in a study determines the correct statistical test for any given situation.

The **analysis of variance (ANOVA)** test compares the **means of ≥3 groups**. It requires a categorical independent variable (ie, exposure) that is used to divide the study pool into ≥3 groups and a quantitative dependent variable (ie, outcome) for which an average (eg, mean) can be calculated. In this study:

- The quantitative dependent variable was the **levels of creatine kinase-MB** (CK-MB).
- The categorical independent variable was type of statin therapy that was used to categorize **4 different groups** of patients with type 2 diabetes mellitus (T2DM): Group I (atorvastatin), Group II (simvastatin), Group III (rosuvastatin), and Group IV (no statin).

An ANOVA test can determine whether there is a statistically significant difference in mean levels of CK-MB among T2DM patients given different statin therapies. A large, statistically significant difference in mean CK-MB levels among groups indicates that different statin therapies are associated with significant changes in CK-MB levels (ie, the null hypothesis is rejected).

(Choice B) The chi-square test evaluates the association between 2 categorical variables, as in a study evaluating the association between sex (ie, "male" and "female") and myocardial infarction (ie, presence of myocardial infarction, absence of myocardial infarction).

Variables are broadly classified as qualitative (ie, categorical) or quantitative (eg, continuous) based on their scale of measurement. **Qualitative variables** (eg, type of treatment, blood type) represent categories or groups, whereas **quantitative variables** (eg, temperature, glucose levels) represent numerical values, with quantitative variables (eg, temperature) sometimes transformed into qualitative variables (eg, "no fever" for <38 C [100 F]; "fever" for ≥38 C [100 F]). The scale of measurement of the dependent (eg, outcome) and independent (eg, exposures, risk factors) variables in a study determines the correct statistical test for any given situation.

The **analysis of variance (ANOVA)** test compares the **means of ≥3 groups**. It requires a categorical independent variable (ie, exposure) that is used to divide the study pool into ≥3 groups and a quantitative dependent variable (ie, outcome) for which an average (eg, mean) can be calculated. In this study:

- The quantitative dependent variable was the **levels of creatine kinase-MB** (CK-MB).
- The categorical independent variable was type of statin therapy that was used to categorize **4 different groups** of patients with type 2 diabetes mellitus (T2DM): Group I (atorvastatin), Group II (simvastatin), Group III (rosuvastatin), and Group IV (no statin).

An ANOVA test can determine whether there is a statistically significant difference in mean levels of CK-MB among T2DM patients given different statin therapies. A large, statistically significant difference in mean CK-MB levels among groups indicates that different statin therapies are associated with significant changes in CK-MB levels (ie, the null hypothesis is rejected).

(Choice B) The chi-square test evaluates the association between 2 categorical variables, as in a study evaluating the association between sex (ie, "male" and "female") and myocardial infarction (ie, presence of myocardial infarction, absence of myocardial infarction).

(Choices C and D) The *t*-tests compare the mean of a quantitative variable *between 2 groups*. A paired *t*-test is used when the groups are dependent (eg, comparing mean blood pressure [quantitative variable] in the same group of patients before and after antihypertensive therapy). An independent *t*-test is used when the groups are independent (eg, comparing mean CK-MB levels [quantitative variable] between a group of patients given statin therapy and another not given statin therapy).

(Choice E) A correlation analysis uses the correlation coefficient to describe the linear relationship between 2 quantitative variables, as in a study evaluating the linear relationship between hours of sleep and irritability score.

Educational objective:
The analysis of variance test compares the means of ≥3 groups. The test requires a categorical independent variable (ie, exposure) that is used

- The quantitative dependent variable was the **levels of creatine kinase-MB** (CK-MB).
- The categorical independent variable was type of statin therapy that was used to categorize **4 different groups** of patients with type 2 diabetes mellitus (T2DM): Group I (atorvastatin), Group II (simvastatin), Group III (rosuvastatin), and Group IV (no statin).

An ANOVA test can determine whether there is a statistically significant difference in mean levels of CK-MB among T2DM patients given different statin therapies. A large, statistically significant difference in mean CK-MB levels among groups indicates that different statin therapies are associated with significant changes in CK-MB levels (ie, the null hypothesis is rejected).

(Choice B) The chi-square test evaluates the association between 2 categorical variables, as in a study evaluating the association between sex (ie, "male" and "female") and myocardial infarction (ie, presence of myocardial infarction, absence of myocardial infarction).

(Choices C and D) The *t*-tests compare the mean of a quantitative variable *between 2 groups*. A paired *t*-test is used when the groups are dependent (eg, comparing mean blood pressure [quantitative variable] in the same group of patients before and after antihypertensive therapy). An independent *t*-test is used when the groups are independent (eg, comparing mean CK-MB levels [quantitative variable] between a group of patients given statin therapy and another not given statin therapy).

(Choice E) A correlation analysis uses the correlation coefficient to describe the linear relationship between 2 quantitative variables, as in a study evaluating the linear relationship between hours of sleep and irritability score.

Educational objective:
The analysis of variance test compares the means of ≥3 groups. The test requires a categorical independent variable (ie, exposure) that is used to divide the study pool into ≥3 groups and a quantitative dependent variable (ie, outcome) for which an average (eg, mean) can be calculated.

References

- Statistical methodology: IV. Analysis of variance, analysis of covariance, and multivariate analysis of variance.

Biostatistics	Biostatistics & Epidemiology	Statistical tests
Subject	System	Topic

Copyright © UWorld. All rights reserved.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Exhibit Display

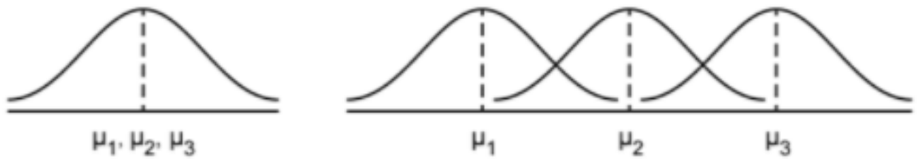
Analysis of variance (ANOVA)

Null hypothesis

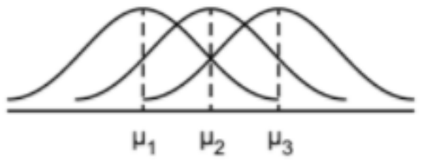
All samples drawn from the same population ($\mu_1=\mu_2=\mu_3$)

Alternate hypothesis

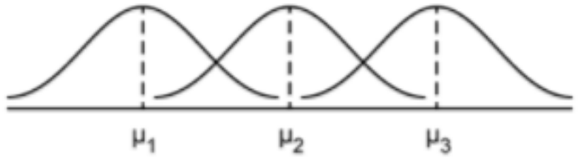
Samples drawn from different populations ($\mu_1\neq\mu_2\neq\mu_3$)



Small apparent differences between the means of the samples: null hypothesis not rejected



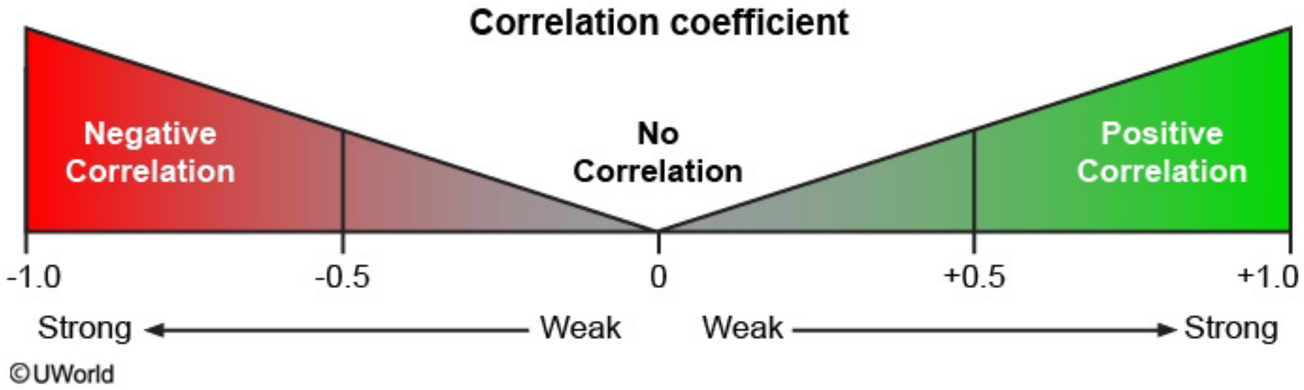
Large apparent differences between the means of the samples: null hypothesis rejected



©UWorld

Zoom In Zoom Out Reset Add To Flash Card

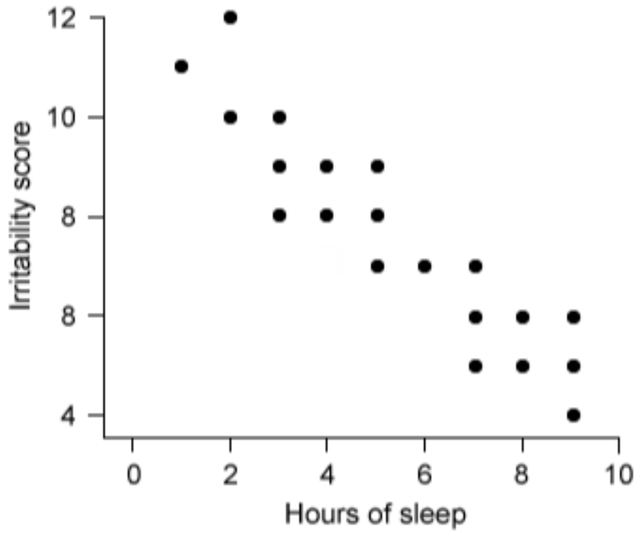
Exhibit Display



Zoom In
Zoom Out
Reset
Add To Flash Card

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Exhibit Display



Zoom In
Zoom Out
Reset
Add To Flash Card

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

A case-control study was conducted to estimate the association between simvastatin therapy and serum levels of fibrinogen in patients who underwent percutaneous coronary intervention. Cases were identified as patients who underwent percutaneous coronary intervention and had high periprocedural levels of fibrinogen (>400 mg/dL), and controls were identified as patients who underwent percutaneous coronary interventions but had normal levels of fibrinogen (200-400 mg/dL). History of simvastatin therapy use was assessed through chart review for every patient. The number of patients corresponding to each classification criteria is given in the table below.

	Fibrinogen high (>400 mg/dL)	Fibrinogen normal (200-400 mg/dL)	Total
Simvastatin therapy	43	67	110
No simvastatin therapy	32	58	90
Total	75	125	200

Which of the following is the best statistical method to estimate the association between simvastatin use and high serum fibrinogen levels in this study?

A. Analysis of variance
B. Chi-square test
C. Correlation analysis
D. Meta-analysis
E. Two-sample *t*-test

Submit

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

A case-control study was conducted to estimate the association between simvastatin therapy and serum levels of fibrinogen in patients who underwent percutaneous coronary intervention. Cases were identified as patients who underwent percutaneous coronary intervention and had high periprocedural levels of fibrinogen (>400 mg/dL), and controls were identified as patients who underwent percutaneous coronary interventions but had normal levels of fibrinogen (200-400 mg/dL). History of simvastatin therapy use was assessed through chart review for every patient. The number of patients corresponding to each classification criteria is given in the table below.

	Fibrinogen high (>400 mg/dL)	Fibrinogen normal (200-400 mg/dL)	Total
Simvastatin therapy	43	67	110
No simvastatin therapy	32	58	90
Total	75	125	200

Which of the following is the best statistical method to estimate the association between simvastatin use and high serum fibrinogen levels in this study?

A. Analysis of variance (4%)
✓ B. Chi-square test (48%)
C. Correlation analysis (17%)
D. Meta-analysis (0%)
E. Two-sample *t*-test (28%)

Omitted
Correct answer
B

48%
Answered correctly

02 secs
Time Spent

03/06/2020
Last Updated

Explanation

		Dependent variable	
		Qualitative (categorical)	*Quantitative*
Independent variable	*Qualitative (categorical)*	Chi-square, logistic regression*	*t*-test, ANOVA, linear regression
	Quantitative	Logistic regression*	Correlation, linear regression

* Dependent variable must be dichotomous.

ANOVA = analysis of variance.

Variables can be broadly classified as qualitative (ie, categorical) or quantitative (eg, continuous) based on their scale of measurement. **Qualitative variables** (eg, disease status, blood type) represent **categories or groups**, whereas **quantitative variables** (eg, body weight, glucose levels) represent **numerical values**. Choosing the correct statistical test always depends on the type of dependent and independent variables under consideration.

The **chi-square test** for independence (also known as chi-square test for association) is used to evaluate the association between **2 categorical variables**. In this example, patients are divided into 2 groups based on serum **fibrinogen levels** (dependent categorical variable: **normal** [200-400 mg/dL], **high** [>400 mg/dL]) and then divided again based on **simvastatin exposure** status (independent categorical variable: **treated**, **not treated**), with the data recorded in a 2 × 2 table. A chi-square test may be conducted to determine whether a statistically significant association exists between simvastatin use and the high levels of serum fibrinogen. A large difference in the frequency of simvastatin use between cases and controls is indicative of an association between simvastatin use and high levels of fibrinogen (ie, the null hypothesis is rejected).

(Choices A and E) The analysis of variance (ANOVA) and two-sample *t*-test are used to compare the *means* between groups. This would require the dependent variable (ie, outcome) to be a *quantitative* variable. The difference between these 2 tests resides in the number of groups compared. The *t*-test is used to compare 2 group means, whereas the ANOVA is used to compare ≥2 group means. Either of these tests would be appropriate in this example if fibrinogen levels were recorded as a quantitative variable (ie, the actual measured level in mg/dL) instead of a categorical variable (ie, normal or high).

(Choice C) The correlation coefficient is a measure of the strength and direction of a linear relationship between 2 *quantitative* variables. For

Variables can be broadly classified as qualitative (ie, categorical) or quantitative (eg, continuous) based on their scale of measurement. **Qualitative variables** (eg, disease status, blood type) represent **categories or groups**, whereas **quantitative variables** (eg, body weight, glucose levels) represent **numerical values**. Choosing the correct statistical test always depends on the type of dependent and independent variables under consideration.

The **chi-square test** for independence (also known as chi-square test for association) is used to evaluate the association between **2 categorical variables**. In this example, patients are divided into 2 groups based on serum **fibrinogen levels** (dependent categorical variable: **normal** [200-400 mg/dL], **high** [>400 mg/dL]) and then divided again based on **simvastatin exposure** status (independent categorical variable: **treated**, **not treated**), with the data recorded in a 2 × 2 table. A chi-square test may be conducted to determine whether a statistically significant association exists between simvastatin use and the high levels of serum fibrinogen. A large difference in the frequency of simvastatin use between cases and controls is indicative of an association between simvastatin use and high levels of fibrinogen (ie, the null hypothesis is rejected).

(Choices A and E) The analysis of variance (ANOVA) and two-sample *t*-test are used to compare the *means* between groups. This would require the dependent variable (ie, outcome) to be a *quantitative* variable. The difference between these 2 tests resides in the number of groups compared. The *t*-test is used to compare 2 group means, whereas the ANOVA is used to compare ≥2 group means. Either of these tests would be appropriate in this example if fibrinogen levels were recorded as a quantitative variable (ie, the actual measured level in mg/dL) instead of a categorical variable (ie, normal or high).

(Choice C) The correlation coefficient is a measure of the strength and direction of a linear relationship between 2 *quantitative* variables. For example, a study may report a correlation coefficient describing the association between hemoglobin A1c levels and average blood glucose levels.

(Choice D) Meta-analysis is an epidemiological method of pooling the data from several studies to conduct an analysis having a relatively larger statistical power than that of the individual studies.

Educational objective:
A 2 × 2 table is normally used to record the presence or absence of exposure and disease in research. Rows and columns represent the different levels for each categorical (ie, exposure and disease) variable. The chi-square test for independence is used to evaluate the association between 2 categorical variables.

Biostatistics Biostatistics & Epidemiology Statistical tests

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

variables under consideration.

The **chi-square test** for independence (also known as chi-square test for association) is used to evaluate the association between **2 categorical variables**. In this example, patients are divided into 2 groups based on serum **fibrinogen levels** (dependent categorical variable: **normal** [200-400 mg/dL], **high** [>400 mg/dL]) and then divided again based on **simvastatin exposure** status (independent categorical variable: **treated**, **not treated**), with the data recorded in a 2 × 2 table. A chi-square test may be conducted to determine whether a statistically significant association exists between simvastatin use and the high levels of serum fibrinogen. A large difference in the frequency of simvastatin use between cases and controls is indicative of an association between simvastatin use and high levels of fibrinogen (ie, the null hypothesis is rejected).

(Choices A and E) The analysis of variance (ANOVA) and two-sample *t*-test are used to compare the *means* between groups. This would require the dependent variable (ie, outcome) to be a *quantitative* variable. The difference between these 2 tests resides in the number of groups compared. The *t*-test is used to compare 2 group means, whereas the ANOVA is used to compare ≥2 group means. Either of these tests would be appropriate in this example if fibrinogen levels were recorded as a quantitative variable (ie, the actual measured level in mg/dL) instead of a categorical variable (ie, normal or high).

(Choice C) The correlation coefficient is a measure of the strength and direction of a linear relationship between 2 *quantitative* variables. For example, a study may report a correlation coefficient describing the association between hemoglobin A1c levels and average blood glucose levels.

(Choice D) Meta-analysis is an epidemiological method of pooling the data from several studies to conduct an analysis having a relatively larger statistical power than that of the individual studies.

Educational objective:
A 2 × 2 table is normally used to record the presence or absence of exposure and disease in research. Rows and columns represent the different levels for each categorical (ie, exposure and disease) variable. The chi-square test for independence is used to evaluate the association between 2 categorical variables.

Biostatistics	Biostatistics & Epidemiology	Statistical tests
Subject	System	Topic

Copyright © UWorld. All rights reserved.

A genome-wide association study (GWAS) of rheumatoid arthritis (RA) is performed in a cohort of Portuguese patients, including 907 cases with RA and 1,524 controls without RA. Logistic regression is used to test the association between RA and hundreds of thousands of loci. These association results are then compared with data from a European GWAS cohort of 4,036 patients with RA and 6,959 patients without RA. Finally, the Portuguese and European study results are combined into a meta-analysis. Based on these data, the investigators identify 3 new loci that are associated with RA at a significance level of 5×10^{-8}:

	European study		Portuguese study		Meta-analysis	
	Odds ratio	*p*-value	Odds ratio	*p*-value	Odds ratio	*p*-value
Locus 1	0.53	8.36×10^{-7}	0.61	0.013	0.55	3.5×10^{-8}
Locus 2	1.16	8.40×10^{-7}	1.14	0.019	1.16	4.9×10^{-8}
Locus 3	1.63	0.0000015	1.77	0.0074	1.66	4.1×10^{-8}

Previous studies had identified 30 loci that accounted for <35% of disease heritability for RA. Which of the following statements is correct regarding this study?

A. Combining the results of the European and Portuguese studies into a meta-analysis decreases the power

B. Logistic regression is used because there is a stepwise increase in the strength of association from Locus 1 to Locus 3

C. The association between Locus 1 and RA in the meta-analysis is not statistically significant because the odds ratio is <1

D. The threshold used for the *p*-value in the meta-analysis is 5×10^{-8} because of the large number of loci studied

E. With this meta-analysis, the 33 identified loci account for most disease heritability in RA

Submit

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

A genome-wide association study (GWAS) of rheumatoid arthritis (RA) is performed in a cohort of Portuguese patients, including 907 cases with RA and 1,524 controls without RA. Logistic regression is used to test the association between RA and hundreds of thousands of loci. These association results are then compared with data from a European GWAS cohort of 4,036 patients with RA and 6,959 patients without RA. Finally, the Portuguese and European study results are combined into a meta-analysis. Based on these data, the investigators identify 3 new loci that are associated with RA at a significance level of 5×10^{-8}:

	European study		Portuguese study		Meta-analysis	
	Odds ratio	*p*-value	Odds ratio	*p*-value	Odds ratio	*p*-value
Locus 1	0.53	8.36×10^{-7}	0.61	0.013	0.55	3.5×10^{-8}
Locus 2	1.16	8.40×10^{-7}	1.14	0.019	1.16	4.9×10^{-8}
Locus 3	1.63	0.0000015	1.77	0.0074	1.66	4.1×10^{-8}

Previous studies had identified 30 loci that accounted for <35% of disease heritability for RA. Which of the following statements is correct regarding this study?

A. Combining the results of the European and Portuguese studies into a meta-analysis decreases the power (5%)

B. Logistic regression is used because there is a stepwise increase in the strength of association from Locus 1 to Locus 3 (26%)

C. The association between Locus 1 and RA in the meta-analysis is not statistically significant because the odds ratio is <1 (19%)

✓ D. The threshold used for the *p*-value in the meta-analysis is 5×10^{-8} because of the large number of loci studied (38%)

E. With this meta-analysis, the 33 identified loci account for most disease heritability in RA (10%)

Omitted
Correct answer

38%

02 secs

06/03/2020

Genome-wide association studies (GWAS) are typically case-control studies that involve scanning thousands of genetic markers in subjects with and without a disease to identify **associations between genetic variants** (eg, loci) and that **disease**. GWAS often use single-nucleotide polymorphisms (SNPs), with several million SNPs analyzed on 1 microscope slide.

In this case, GWAS are used to analyze **hundreds of thousands of loci** in Portuguese patients with and without rheumatoid arthritis (RA). The results are compared to those from a larger European cohort; both studies are then combined in a meta-analysis. A few loci have SNPs with an allele that is significantly more common among patients with RA than those without RA; these loci are therefore associated with RA.

Whenever **multiple tests** are performed simultaneously (eg, analyzing thousands of loci), the possibility of **false-positive results** increases. Using a much **smaller threshold for the *p*-value** is a method to minimize this. Traditionally, a *p*-value <5% is considered statistically significant. However, if 100,000 loci are studied at that 5% (ie, 0.05, 5×10^{-2}) level, then 100,000 × 0.05 = 5,000 false positives would be expected. With a much smaller *p*-value, this number decreases substantially. For this reason, a *p*-value of 5×10^{-8} ("genome-wide *p*-value"), as seen here, is often chosen in GWAS.

(Choice A) A meta-analysis combines results from different studies, thereby increasing sample size, which increases power (eg, to detect statistical differences between groups).

(Choice B) Logistic regression was used because it analyzes the association between exposures (eg, multiple loci) and binary outcomes (eg, yes = cases with RA; no = controls without RA).

(Choice C) An odds ratio (OR) <1 does not imply a lack of statistically significant association. If an OR >1 means that patients with RA have *higher* odds of having a specific locus compared to those without RA (ie, risk factor), an OR <1 means that patients with RA have *lower* odds of having that specific locus (ie, protective factor). The OR for the association between Locus 1 and RA is statistically significant because its *p*-value (3.4×10^{-8}) is less than the genome-wide *p*-value (5×10^{-8}).

(Choice E) GWAS help explain disease heritability (ie, effect of genotypic differences on phenotypic differences). Although this study identified 3 new loci, there is no indication that the full group of 33 loci will explain most RA disease heritability. GWAS have identified many genes associated with specific conditions, but these genes only seem to explain a small amount of the variance ("missing heritability").

Educational objective:

Genome-wide association studies aim to identify associations between thousands of genetic variants and a disease. Because of the increased

Using a much **smaller threshold for the *p*-value** is a method to minimize this. Traditionally, a *p*-value <5% is considered statistically significant. However, if 100,000 loci are studied at that 5% (ie, 0.05, 5×10^{-2}) level, then 100,000 × 0.05 = 5,000 false positives would be expected. With a much smaller *p*-value, this number decreases substantially. For this reason, a *p*-value of 5×10^{-8} ("genome-wide *p*-value"), as seen here, is often chosen in GWAS.

(Choice A) A meta-analysis combines results from different studies, thereby increasing sample size, which increases power (eg, to detect statistical differences between groups).

(Choice B) Logistic regression was used because it analyzes the association between exposures (eg, multiple loci) and binary outcomes (eg, yes = cases with RA; no = controls without RA).

(Choice C) An odds ratio (OR) <1 does not imply a lack of statistically significant association. If an OR >1 means that patients with RA have *higher* odds of having a specific locus compared to those without RA (ie, risk factor), an OR <1 means that patients with RA have *lower* odds of having that specific locus (ie, protective factor). The OR for the association between Locus 1 and RA is statistically significant because its *p*-value (3.4×10^{-8}) is less than the genome-wide *p*-value (5×10^{-8}).

(Choice E) GWAS help explain disease heritability (ie, effect of genotypic differences on phenotypic differences). Although this study identified 3 new loci, there is no indication that the full group of 33 loci will explain most RA disease heritability. GWAS have identified many genes associated with specific conditions, but these genes only seem to explain a small amount of the variance ("missing heritability").

Educational objective:
Genome-wide association studies aim to identify associations between thousands of genetic variants and a disease. Because of the increased risk of false-positive results when multiple tests are performed simultaneously, a smaller genome-wide *p*-value is typically used.

References

- GWAS identifies novel SLE susceptibility genes and explains the association of the HLA region.

Biostatistics	Biostatistics & Epidemiology	P-value and confidence interval
Subject	System	Topic

Copyright © UWorld. All rights reserved.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

An intern in the endocrinology department is asked to complete a chart review of cholesterol levels of all patients hospitalized with diabetes mellitus-related cardiovascular complications in the past 6 months. The intern's preliminary data analysis reveals that there were 400 such patients, and that serum cholesterol levels were normally distributed among the patients with a mean of 220 mg/dL and a standard deviation of 10 mg/dL. Based on these results, how many patients in this study would be expected to have serum cholesterol ≥240 mg/dL?

A. 2

B. 10

C. 20

D. 64

E. 128

Submit

An intern in the endocrinology department is asked to complete a chart review of cholesterol levels of all patients hospitalized with diabetes mellitus-related cardiovascular complications in the past 6 months. The intern's preliminary data analysis reveals that there were 400 such patients, and that serum cholesterol levels were normally distributed among the patients with a mean of 220 mg/dL and a standard deviation of 10 mg/dL. Based on these results, how many patients in this study would be expected to have serum cholesterol ≥240 mg/dL?

- A. 2 (9%)
- ✓ B. 10 (65%)
- C. 20 (18%)
- D. 64 (4%)
- E. 128 (1%)

Omitted
Correct answer
B

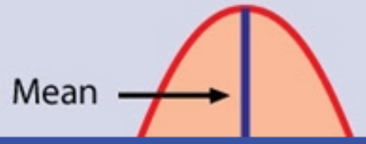
65%
Answered correctly

02 secs
Time Spent

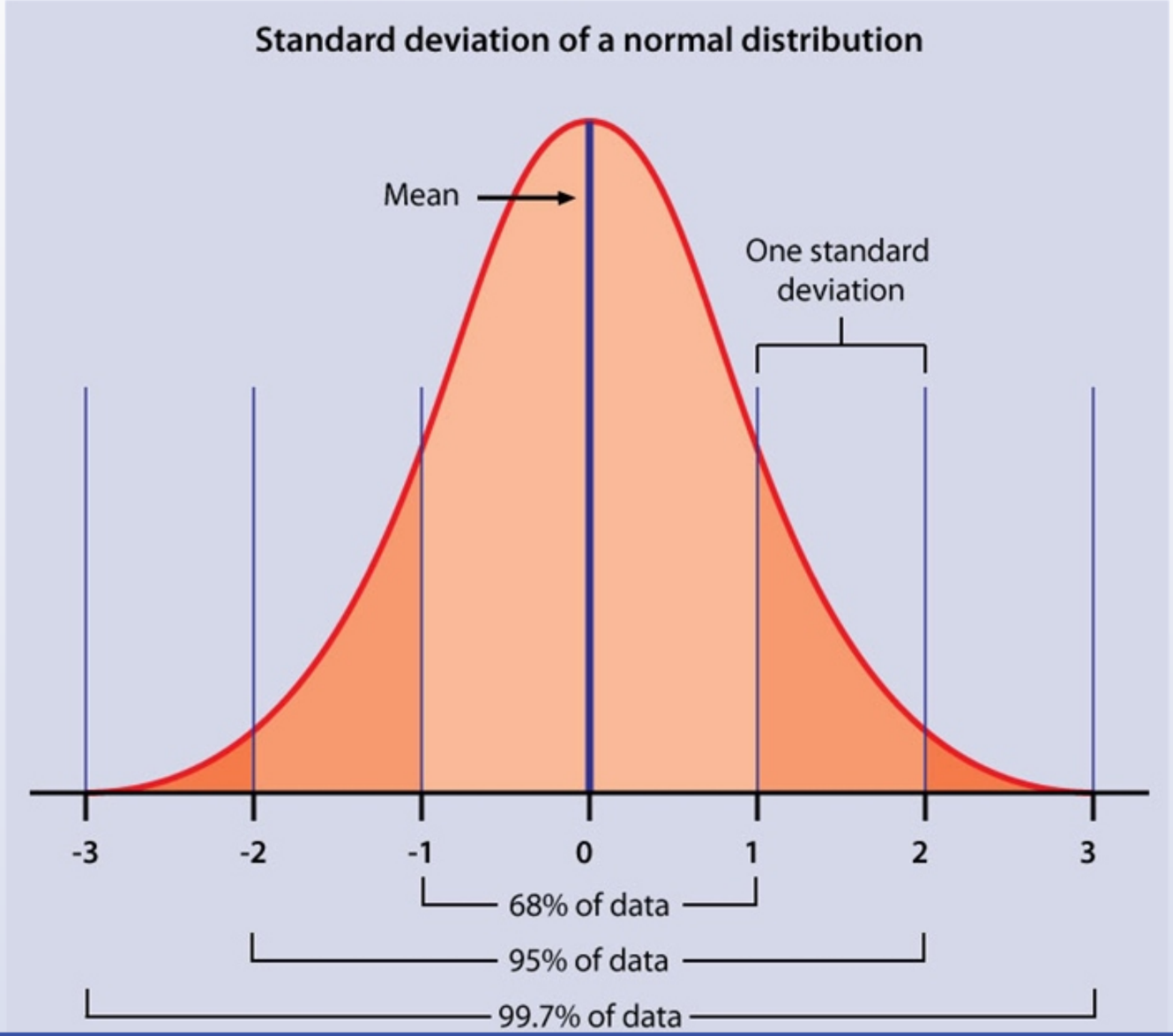
05/12/2020
Last Updated

Explanation

Standard deviation of a normal distribution



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31



A **normal distribution** refers to a symmetrical, bell-shaped distribution pattern with a fixed proportion of observations lying within a certain distance of the mean. This distance is called the **standard deviation (SD)** and is a measure of the **degree of dispersion** from the mean. When the SD is small, the observations tend to have minimal variation and are tightly clustered around the mean. In contrast, a large SD implies that the observations are spread over a large range. According to the properties of normal distribution, 68% of all observations lie within 1 SD of the mean, with half of those above and half of those below the mean. The remaining 32% of observations lie outside 1 SD from the mean, with half of those remaining observations above and the other half below 1 SD from the mean. In addition, **95% of all observations lie within 2 SDs** of the mean, and 99.7% of all observations lie within 3 SDs from the mean. This is commonly known as the "**68-95-99.7 rule**."

Based on the above example:

- 68% of observations lie within 1 SD: 220 mg/dL ± 10 mg/dL = 210-230 mg/dL
- 95% of observations lie within 2 SDs: 220 mg/dL ± 20 mg/dL = 200-240 mg/dL
- 99.7% of observations lie within 3 SDs: 220 mg/dL ± 30 mg/dL = 190-250 mg/dL

Because the cutoff point of ≥240 mg/dL is 2 SDs from the mean, it follows that 2.5% of the observations must lie below 200 mg/dL and **2.5% above 240 mg/dL**. Therefore, of 400 patients included in the analysis, **10 patients** (ie, 2.5%) would have a cholesterol level ≥240 mg/dL.

(Choice A) No more than 1-2 patients (0.25%-0.5% of the study population) would be expected to have cholesterol levels that lie 3 SDs above or below the mean (ie, ≤190 mg/dL or ≥250 mg/dL).

(Choice C) 20 patients (5% of the study population) would be expected to have cholesterol levels that lie outside 2 SDs from the mean. Only half of these observations would lie above 2 SDs (ie, ≥240 mg/dL) from the mean.

(Choice D) 64 patients (16% of the study population) would be expected to have cholesterol levels that lie 1 SD above the mean (ie, ≥230 mg/dL).

(Choice E) 128 patients (32% of the study population) would be expected to have cholesterol levels that lie 1 SD above or below the mean (ie, ≥230 mg/dL or ≤210 mg/dL).

Educational objective:
In a normal (bell-shaped) distribution: 68% of all values are within 1 standard deviation from the mean; 95% of all values are within 2 standard deviations from the mean; 99.7% of all values are within 3 standard deviations from the mean.

those remaining observations above and the other half below 1 SD from the mean. In addition, **95% of all observations lie within 2 SDs** of the mean, and 99.7% of all observations lie within 3 SDs from the mean. This is commonly known as the "**68-95-99.7 rule**."

Based on the above example:

- 68% of observations lie within 1 SD: 220 mg/dL ± 10 mg/dL = 210-230 mg/dL
- 95% of observations lie within 2 SDs: 220 mg/dL ± 20 mg/dL = 200-240 mg/dL
- 99.7% of observations lie within 3 SDs: 220 mg/dL ± 30 mg/dL = 190-250 mg/dL

Because the cutoff point of ≥240 mg/dL is 2 SDs from the mean, it follows that 2.5% of the observations must lie below 200 mg/dL and **2.5% above 240 mg/dL**. Therefore, of 400 patients included in the analysis, **10 patients** (ie, 2.5%) would have a cholesterol level ≥240 mg/dL.

(Choice A) No more than 1-2 patients (0.25%-0.5% of the study population) would be expected to have cholesterol levels that lie 3 SDs above or below the mean (ie, ≤190 mg/dL or ≥250 mg/dL).

(Choice C) 20 patients (5% of the study population) would be expected to have cholesterol levels that lie outside 2 SDs from the mean. Only half of these observations would lie above 2 SDs (ie, ≥240 mg/dL) from the mean.

(Choice D) 64 patients (16% of the study population) would be expected to have cholesterol levels that lie 1 SD above the mean (ie, ≥230 mg/dL).

(Choice E) 128 patients (32% of the study population) would be expected to have cholesterol levels that lie 1 SD above or below the mean (ie, ≥230 mg/dL or ≤210 mg/dL).

Educational objective:
In a normal (bell-shaped) distribution: 68% of all values are within 1 standard deviation from the mean; 95% of all values are within 2 standard deviations from the mean; 99.7% of all values are within 3 standard deviations from the mean.

Biostatistics	Biostatistics & Epidemiology	Normal distribution
Subject	System	Topic

Copyright © UWorld. All rights reserved.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

A geneticist is interested in the potential causes of a congenital abnormality. She hypothesizes that acetaminophen use during the first trimester of pregnancy might be associated with the abnormality. In her study, a sample of mothers of children with and without the abnormality are randomly selected from government birth records. Personal interviews are then conducted with the mothers to determine fetal exposure to acetaminophen. Results show that mothers of children who do not have the abnormality did not take acetaminophen as frequently during the first trimester. This type of study is most susceptible to which of the following types of bias?

A. Allocation bias

B. Detection bias

C. Recall bias

D. Referral bias

E. Selection bias

Submit

A geneticist is interested in the potential causes of a congenital abnormality. She hypothesizes that acetaminophen use during the first trimester of pregnancy might be associated with the abnormality. In her study, a sample of mothers of children with and without the abnormality are randomly selected from government birth records. Personal interviews are then conducted with the mothers to determine fetal exposure to acetaminophen. Results show that mothers of children who do not have the abnormality did not take acetaminophen as frequently during the first trimester. This type of study is most susceptible to which of the following types of bias?

A. Allocation bias (0%)

B. Detection bias (2%)

✓ C. Recall bias (92%)

D. Referral bias (0%)

E. Selection bias (4%)

Omitted
Correct answer
C

92%
Answered correctly

02 secs
Time Spent

01/16/2020
Last Updated

Explanation

Recall bias results from study participants' **inaccurate recall** of past exposure and occurs most often in **retrospective** studies such as **case-control** studies. People who have experienced an adverse event (eg, congenital abnormalities in their children) are more likely to recall previous potential risk factors (eg, acetaminophen use) than people who have not experienced an adverse event.

Contrary to case-control studies, which evaluate exposure to potential risk factors retrospectively, **prospective studies** begin with exposed and unexposed individuals. Because the exposure status is determined at the time of enrollment in the study, recall bias is **minimized**.

Explanation

Recall bias results from study participants' **inaccurate recall** of past exposure and occurs most often in **retrospective** studies such as **case-control** studies. People who have experienced an adverse event (eg, congenital abnormalities in their children) are more likely to recall previous potential risk factors (eg, acetaminophen use) than people who have not experienced an adverse event.

Contrary to case-control studies, which evaluate exposure to potential risk factors retrospectively, **prospective studies** begin with exposed and unexposed individuals. Because the exposure status is determined at the time of enrollment in the study, recall bias is **minimized**.

(Choice A) Allocation bias can result from the way patients are *assigned* to the treatment and control groups. It may occur when subjects are nonrandomly assigned to the study groups of a clinical trial (eg, physicians may preferentially enroll sicker patients into the experimental group). Allocation bias is different from selection bias, which occurs when the studied sample does not represent the general population (eg, because of *nonrandom selection*).

(Choice B) Detection bias refers to the fact that a risk factor itself may lead to extensive diagnostic investigation and increase the probability that a disease is identified. For instance, patients who smoke may undergo increased imaging surveillance due to their smoking status, which would detect more cases of cancer in general.

(Choice D) Referral (admission rate) bias occurs when the case and control populations differ due to admission or referral practices. For instance, a study involving asbestos as a risk factor for lung cancer conducted at a hospital specializing in treating asbestosis may select patients with lung cancer (cases) from the respiratory department and a control group without lung cancer from other departments. Because the hospital specializes in asbestosis treatment, patients in the respiratory department with lung cancer are more likely to have a history of asbestos exposure (compared to patients with lung cancer at other, nonspecialized institutions). Therefore, the study may erroneously report a stronger association than actually exists.

(Choice E) Selection bias occurs when the method of selection used in a particular study results in selection sample of participants who are not representative of the intended population. A common example is patients with cancer who fail standard therapy and who are therefore the most likely to enroll in experimental trials (self-selection), leading to results that are not applicable to patients with less advanced cancers.

Educational objective:
Recall bias results from study participants' inaccurate recall of past exposure and occurs most often in retrospective studies such as case-control studies. People who have experienced an adverse event are more likely to recall risk factors than those who have not experienced an adverse

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

(Choice A) Allocation bias can result from the way patients are *assigned* to the treatment and control groups. It may occur when subjects are nonrandomly assigned to the study groups of a clinical trial (eg, physicians may preferentially enroll sicker patients into the experimental group). Allocation bias is different from selection bias, which occurs when the studied sample does not represent the general population (eg, because of *nonrandom selection*).

(Choice B) Detection bias refers to the fact that a risk factor itself may lead to extensive diagnostic investigation and increase the probability that a disease is identified. For instance, patients who smoke may undergo increased imaging surveillance due to their smoking status, which would detect more cases of cancer in general.

(Choice D) Referral (admission rate) bias occurs when the case and control populations differ due to admission or referral practices. For instance, a study involving asbestos as a risk factor for lung cancer conducted at a hospital specializing in treating asbestosis may select patients with lung cancer (cases) from the respiratory department and a control group without lung cancer from other departments. Because the hospital specializes in asbestosis treatment, patients in the respiratory department with lung cancer are more likely to have a history of asbestos exposure (compared to patients with lung cancer at other, nonspecialized institutions). Therefore, the study may erroneously report a stronger association than actually exists.

(Choice E) Selection bias occurs when the method of selection used in a particular study results in selection sample of participants who are not representative of the intended population. A common example is patients with cancer who fail standard therapy and who are therefore the most likely to enroll in experimental trials (self-selection), leading to results that are not applicable to patients with less advanced cancers.

Educational objective:
Recall bias results from study participants' inaccurate recall of past exposure and occurs most often in retrospective studies such as case-control studies. People who have experienced an adverse event are more likely to recall risk factors than those who have not experienced an adverse event.

Biostatistics	Biostatistics & Epidemiology	Bias
Subject	System	Topic

Copyright © UWorld. All rights reserved.

Potato salad only	10	3
Hamburgers and potato salad	25	5
Hotdogs and potato salad	8	3
Hamburgers, hotdogs, and potato salad	30	14

Which of the following best describes the attack rate among all of the attendees who had potato salad?

- A. 8%
- B. 11%
- C. 13%
- D. 20%
- E. 28%
- F. 30%
- G. 34%
- H. 47%

Submit

A residency program organized a barbecue at the beginning of the academic year to welcome the new class of first-year residents. The total number of attendees was 100, including faculty, administrative staff, and residents. Hamburgers, hotdogs, and potato salad were served. The following day, 28 of the attendees had diarrhea and vomiting. All of the attendees were questioned about what they had eaten, and the following table was obtained:

Food item or combination of items	Number of attendees who ate food item or combination of items	Number of attendees who developed diarrhea and vomiting
Hamburgers only	15	2
Hotdogs only	12	1
Potato salad only	10	3
Hamburgers and potato salad	25	5
Hotdogs and potato salad	8	3
Hamburgers, hotdogs, and potato salad	30	14

Which of the following best describes the attack rate among all of the attendees who had potato salad?

A. 8%

B. 11%

C. 13%

Potato salad only	10	3
Hamburgers and potato salad	25	5
Hotdogs and potato salad	8	3
Hamburgers, hotdogs, and potato salad	30	14

Which of the following best describes the attack rate among all of the attendees who had potato salad?

- A. 8%
- B. 11%
- C. 13%
- D. 20%
- E. 28%
- F. 30%
- G. 34%
- H. 47%

Submit

Potato salad only	10	3
Hamburgers and potato salad	25	5
Hotdogs and potato salad	8	3
Hamburgers, hotdogs, and potato salad	30	14

Which of the following best describes the attack rate among all of the attendees who had potato salad?

- A. 8% (0%)
- B. 11% (1%)
- C. 13% (1%)
- D. 20% (2%)
- E. 28% (5%)
- F. 30% (15%)
- ✓ G. 34% (69%)
- H. 47% (2%)

Omitted
Correct answer

69%
Answered correctly

05 secs
Time Spent

07/04/2020
Last Updated

There appears to be an **outbreak** as almost a third of the attendees (28/100) developed diarrhea and vomiting; this is clearly a larger number than expected. The **attack rate** is the ratio of the number of individuals who become ill divided by the number of individuals who are at risk of contracting that illness; it is often used in outbreak investigations. The attack rate for potato salad corresponds to the proportion of attendees who became ill out of the total number of attendees who had potato salad (alone or in combination with other food items).

Total number of attendees who had potato salad
= Number of attendees who had potato salad only + number of attendees who had hamburgers and potato salad + number of attendees who had hotdogs and potato salad + number of attendees who had hamburgers, hotdogs, and potato salad
= 10 + 25 + 8 + 30
= 73

Among those, the number who became ill was 25 (3 out of the 10 who had potato salad only + 5 out of the 25 who had hamburgers and potato salad + 3 out of the 8 who had hotdogs and potato salad + 14 out of the 30 who had hamburgers, hot dogs, and potato salad = 25). Therefore, the attack rate for potato salad was: (25/73) × 100 = 34%.

(Choice B) There were 27 attendees who did not eat potato salad: the 15 who had hamburgers only and the 12 who had hotdogs only. Of those, a total of 3 (2 out of the 15 who ate hamburgers only + 1 out of the 12 who ate hotdogs only = 3) became ill. Therefore, the attack rate among those who did not eat potato salad was: (3/27) × 100 = 11%. Despite the fact that these individuals only ate hotdogs and hamburgers, a percentage of them still became ill, probably as a result of some contamination of these food items with the potato salad.

For the sake of completion, the attack rates for the other individual food items or combinations of food items are shown in the following table (they do not need to be calculated to arrive at the correct answer):

	Food item or combination of items	Number of attendees who ate food item or combination of items	Number of attendees who developed diarrhea and vomiting	Attack rate
Did not have	Hamburgers			(2/15) × 100 =

	Food item or combination of items	Number of attendees who ate food item or combination of items	Number of attendees who developed diarrhea and vomiting	Attack rate
Did not have potato salad = 27 (attack rate 11%)	Hamburgers only	15	2	(2/15) × 100 = 13% **(Choice C)**
	Hotdogs only	12	1	(1/12) × 100 = 8% **(Choice A)**
Had potato salad = 73 (attack rate 34%)	Potato salad only	10	3	(3/10) × 100 = 30% **(Choice F)**
	Hamburgers and potato salad	25	5	(5/25) × 100 = 20% **(Choice D)**
	Hotdogs and potato salad	8	3	(3/8) × 100 = 38%
	Hamburgers, hotdogs, and potato salad	30	14	(14/30) × 100 = 47% **(Choice H)**
Overall		100	28	28/100 × 100 = 28% **(Choice E)**

salad = 27 (attack rate 11%)	Hotdogs only	12	1	(1/12) × 100 = 8% **(Choice A)**
Had potato salad = 73 (attack rate 34%)	Potato salad only	10	3	(3/10) × 100 = 30% **(Choice F)**
	Hamburgers and potato salad	25	5	(5/25) × 100 = 20% **(Choice D)**
	Hotdogs and potato salad	8	3	(3/8) × 100 = 38%
	Hamburgers, hotdogs, and potato salad	30	14	(14/30) × 100 = 47% **(Choice H)**
Overall		100	28	28/100 × 100 = 28% **(Choice E)**

The food item responsible for an outbreak often (though not always) has the largest attack rate and the largest difference in attack rates between those who consumed the item and those who did not. In this example, as calculated above, the difference in attack rate between those who had potato salad and those who did not was: 34% − 11% = 23%. As seen in the exhibit, the differences in attack rates were smaller for hamburgers (7%) and hotdogs (16%). It is possible that potato salad is the source of the outbreak.

Educational objective:
The attack rate is the ratio of the number of people who contract an illness divided by the number of people who are at risk of contracting that illness.

(attack | Hotdogs only | 12 | 1 | (1/12) × 100 =

Exhibit Display

Calculation of attack rates for different food items in an outbreak at a residency program barbecue

Food item	Attack rate among those who consumed food item	Attack rate among those who did not consume food item	Difference in attack rate
Potato salad†	$\frac{3+5+3+14}{10+25+8+30} \times 100 = 34\%$	$\frac{2+1}{15+12} \times 100 = 11\%$	34% − 11% = 23%
Hamburgers	$\frac{2+5+14}{15+25+30} \times 100 = 30\%$	$\frac{1+3+3}{12+10+8} \times 100 = 23\%$	30% − 23% = 7%
Hotdogs	$\frac{1+3+14}{12+8+30} \times 100 = 36\%$	$\frac{2+3+5}{15+10+25} \times 100 = 20\%$	36% − 20% = 16%

†See calculations for potato salad in text. Similar calculations can be performed for hamburgers & hotdogs.

©UWorld

Zoom In | Zoom Out | Reset | Add To New Card | Existing Card

Biostatistics | Biostatistics & Epidemiology | Epidemics and pandemics

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

A pediatric cardiologist decides to determine whether there is a relationship between body composition and blood pressure in a random sample of children and adolescents age 5-15. He used standard procedures to collect anthropometric measurements that included weight, height, hip circumference (HC), and waist circumference (WC). Systolic blood pressure (SBP) and diastolic pressure (DBP) readings were taken at least 3 times at 5-minute intervals after the participants had been seated, and average SBP and DBP readings were calculated based on these measurements. Which of the following statistical tests is adequate to determine whether there is a relationship between WC and average DBP?

A. Analysis of variance

B. Chi-square test

C. Correlation analysis

D. Meta-analysis

E. Two-sample *t*-test

Submit

A pediatric cardiologist decides to determine whether there is a relationship between body composition and blood pressure in a random sample of children and adolescents age 5-15. He used standard procedures to collect anthropometric measurements that included weight, height, hip circumference (HC), and waist circumference (WC). Systolic blood pressure (SBP) and diastolic pressure (DBP) readings were taken at least 3 times at 5-minute intervals after the participants had been seated, and average SBP and DBP readings were calculated based on these measurements. Which of the following statistical tests is adequate to determine whether there is a relationship between WC and average DBP?

- A. Analysis of variance (5%)
- B. Chi-square test (9%)
- ✓ C. Correlation analysis (55%)
- D. Meta-analysis (1%)
- E. Two-sample *t*-test (27%)

Omitted
Correct answer
C

55%
Answered correctly

02 secs
Time Spent

05/06/2020
Last Updated

Explanation

		Dependent variable	
		Qualitative (categorical)	*Quantitative*
Independent variable	*Qualitative (categorical)*	Chi-square, logistic regression*	*t*-test, ANOVA, linear regression

		Dependent variable	
		Qualitative (categorical)	*Quantitative*
Independent variable	*Qualitative (categorical)*	Chi-square, logistic regression*	*t*-test, ANOVA, linear regression
	Quantitative	Logistic regression*	Correlation, linear regression

* Dependent variable must be dichotomous.

ANOVA = analysis of variance.

Variables are broadly classified as qualitative (ie, categorical) or quantitative (ie, continuous) based on their scale of measurement. **Qualitative variables** (eg, type of treatment, blood type) represent categories or groups whereas **quantitative variables** (eg, temperature, glucose levels) represent numerical values. The scale of measurement of the dependent (eg, outcome) and independent (eg, exposures, risk factors) variables in a study determines the correct statistical test for any given situation.

Correlation analysis is a statistical technique **used to assess** the strength and direction of a **linear relationship between 2 quantitative variables** (often, but not always, a dependent and an independent variable). In this study:

- The quantitative dependent variable was diastolic pressure (DBP).
- The quantitative independent variable was waist circumference (WC).

Correlation analysis can determine whether there is a relationship between DBP and WC. A statistically significant correlation coefficient indicates that DBP and WC have a linear relationship (ie, the null hypothesis is rejected) that is either positive (as WC increases, DBP increases) or negative (as WC increases, DBP decreases).

(Choice A) The analysis of variance (ANOVA) test compares the mean of a quantitative variable of ≥3 independent groups, as in a study comparing serum ferritin concentrations (ie, quantitative variable) in children (age 0-12), adolescents (age 13-17), adults (age 18-59), and seniors (age ≥60).

(Choice B) The chi-square test evaluates the association between 2 categorical variables, as in a study evaluating the association between sex (ie, "male" and "female") and myocardial infarction (ie, presence of myocardial infarction, absence of myocardial infarction).

Variables are broadly classified as qualitative (ie, categorical) or quantitative (ie, continuous) based on their scale of measurement. **Qualitative variables** (eg, type of treatment, blood type) represent categories or groups whereas **quantitative variables** (eg, temperature, glucose levels) represent numerical values. The scale of measurement of the dependent (eg, outcome) and independent (eg, exposures, risk factors) variables in a study determines the correct statistical test for any given situation.

Correlation analysis is a statistical technique **used to assess** the strength and direction of a **linear relationship between 2 quantitative variables** (often, but not always, a dependent and an independent variable). In this study:

- The quantitative dependent variable was diastolic pressure (DBP).
- The quantitative independent variable was waist circumference (WC).

Correlation analysis can determine whether there is a relationship between DBP and WC. A statistically significant correlation coefficient indicates that DBP and WC have a linear relationship (ie, the null hypothesis is rejected) that is either positive (as WC increases, DBP increases) or negative (as WC increases, DBP decreases).

(Choice A) The analysis of variance (ANOVA) test compares the mean of a quantitative variable of ≥3 independent groups, as in a study comparing serum ferritin concentrations (ie, quantitative variable) in children (age 0-12), adolescents (age 13-17), adults (age 18-59), and seniors (age ≥60).

(Choice B) The chi-square test evaluates the association between 2 categorical variables, as in a study evaluating the association between sex (ie, "male" and "female") and myocardial infarction (ie, presence of myocardial infarction, absence of myocardial infarction).

(Choice D) Meta-analysis is a quantitative statistical technique used to combine and analyze data from several studies to conduct an analysis with a greater statistical power than that of the individual studies.

(Choice E) The two-sample *t*-test compares the mean of a quantitative variable between 2 groups, as in a study comparing serum ferritin concentrations (ie, quantitative variable) in male and female patients.

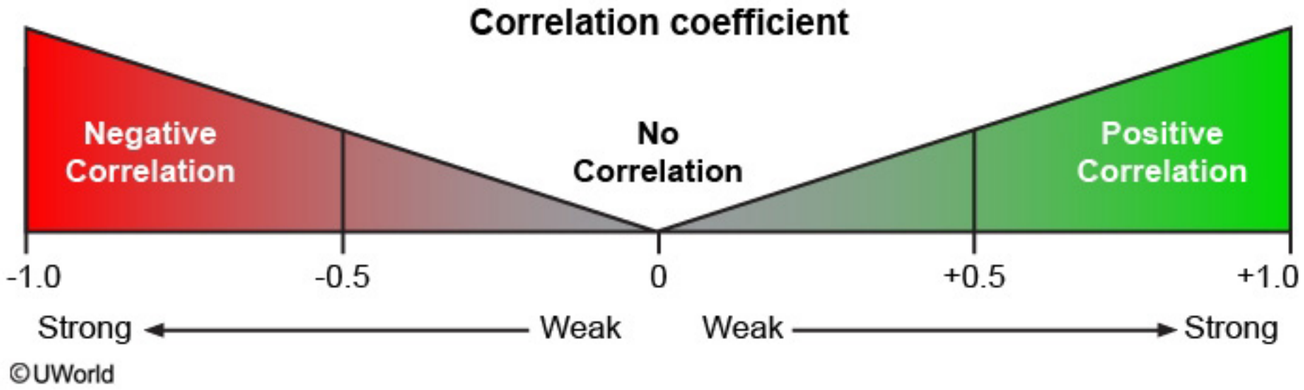
Educational objective:
Correlation analysis is a statistical technique used to describe the strength and direction of a linear relationship between 2 quantitative variables.

References

- Correlation coefficients: appropriate use and interpretation

Variables are broadly classified as qualitative (ie, categorical) or quantitative (ie, continuous) based on their scale of measurement. **Qualitative**

Exhibit Display



Zoom In Zoom Out Reset Add To Flash Card

Correlation coefficients: appropriate use and interpretation

variables (often, but not always, a dependent and an independent variable). In this study:

- The quantitative dependent variable was diastolic pressure (DBP).
- The quantitative independent variable was waist circumference (WC).

Correlation analysis can determine whether there is a relationship between DBP and WC. A statistically significant correlation coefficient indicates that DBP and WC have a linear relationship (ie, the null hypothesis is rejected) that is either positive (as WC increases, DBP increases) or negative (as WC increases, DBP decreases).

(Choice A) The analysis of variance (ANOVA) test compares the mean of a quantitative variable of ≥3 independent groups, as in a study comparing serum ferritin concentrations (ie, quantitative variable) in children (age 0-12), adolescents (age 13-17), adults (age 18-59), and seniors (age ≥60).

(Choice B) The chi-square test evaluates the association between 2 categorical variables, as in a study evaluating the association between sex (ie, "male" and "female") and myocardial infarction (ie, presence of myocardial infarction, absence of myocardial infarction).

(Choice D) Meta-analysis is a quantitative statistical technique used to combine and analyze data from several studies to conduct an analysis with a greater statistical power than that of the individual studies.

(Choice E) The two-sample *t*-test compares the mean of a quantitative variable between 2 groups, as in a study comparing serum ferritin concentrations (ie, quantitative variable) in male and female patients.

Educational objective:
Correlation analysis is a statistical technique used to describe the strength and direction of a linear relationship between 2 quantitative variables.

References

- Correlation coefficients: appropriate use and interpretation.

Biostatistics	Biostatistics & Epidemiology	Statistical tests
Subject	System	Topic

Copyright © UWorld. All rights reserved.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

A study was conducted to assess the age at menarche among young female gymnasts. Based on commitment to the sport, gymnasts were divided into two groups: competitive gymnasts and recreational gymnasts.

Age at menarche

	Sample size (n)	Mean, y	Standard deviation, y
Competitive	16	13.4	1.3
Recreational	22	12.3	0.8

Assuming that age at menarche is normally distributed, which of the following is closest to the probability that a randomly chosen competitive gymnast will have onset of menarche at age ≥16?

- A. 0.997
- B. 0.950
- C. 0.680
- D. 0.160
- E. 0.025
- F. 0.0015

Submit

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

A study was conducted to assess the age at menarche among young female gymnasts. Based on commitment to the sport, gymnasts were divided into two groups: competitive gymnasts and recreational gymnasts.

Age at menarche

	Sample size (n)	Mean, y	Standard deviation, y
Competitive	16	13.4	1.3
Recreational	22	12.3	0.8

Assuming that age at menarche is normally distributed, which of the following is closest to the probability that a randomly chosen competitive gymnast will have onset of menarche at age ≥16?

A. 0.997 (3%)
B. 0.950 (7%)
C. 0.680 (3%)
D. 0.160 (4%)
✓ E. 0.025 (72%)
F. 0.0015 (9%)

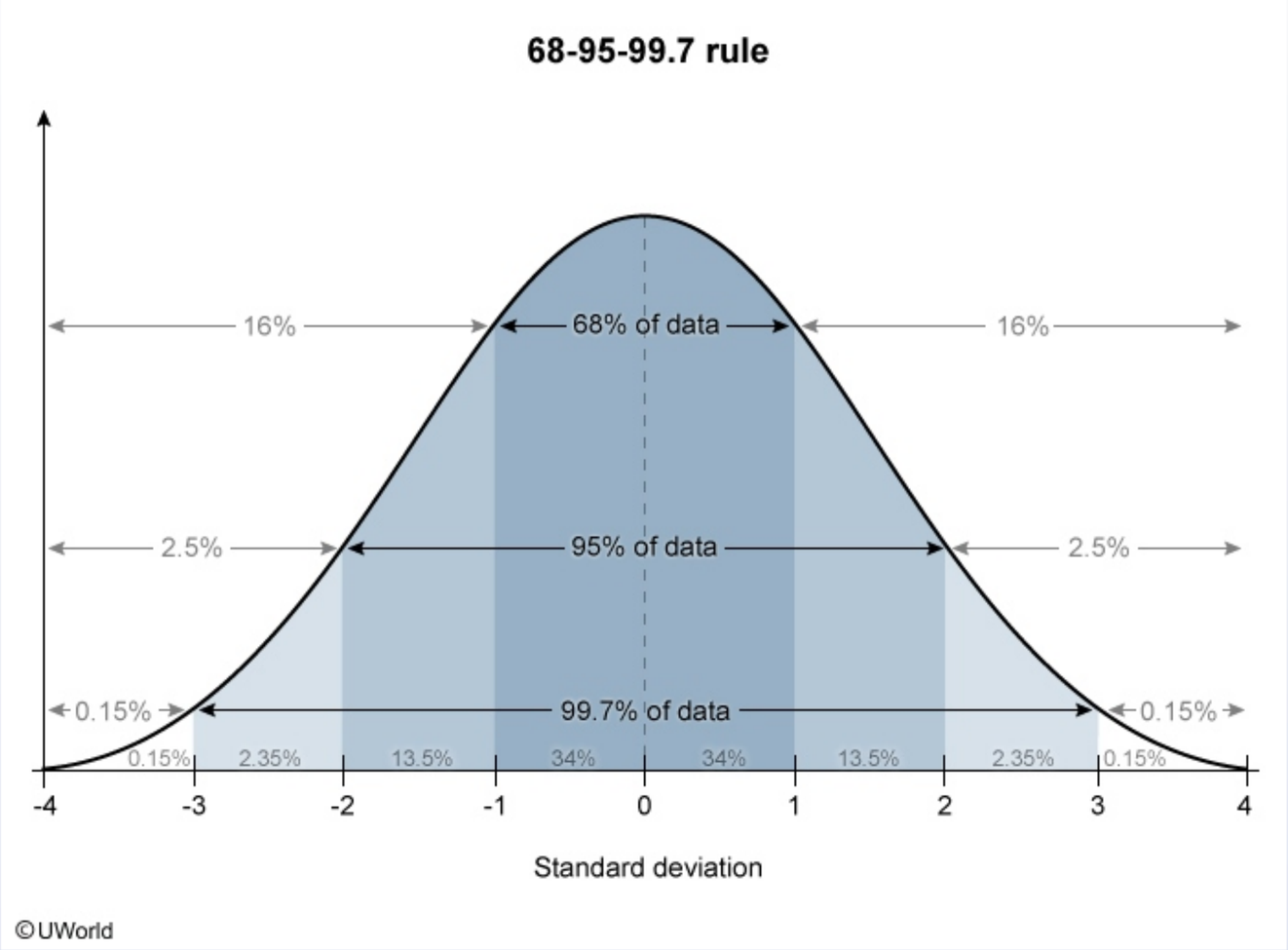
Omitted
Correct answer
E

72%
Answered correctly

02 secs
Time Spent

05/14/2020
Last Updated

Explanation



A normal (Gaussian) distribution is a symmetrical, bell-shaped distribution with a fixed percentage of observations lying within a certain distance of the mean. This distance is called the standard deviation (SD) and represents the degree of dispersion from the mean. The **68-95-99.7** rule for normal distributions states that 68% of all observations lie within 1 SD of the mean, 95% of all observations lie within 2 SDs of the mean, and

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

A normal (Gaussian) distribution is a symmetrical, bell-shaped distribution with a fixed percentage of observations lying within a certain distance of the mean. This distance is called the standard deviation (SD) and represents the degree of dispersion from the mean. The **68-95-99.7** rule for normal distributions states that 68% of all observations lie within 1 SD of the mean, 95% of all observations lie within 2 SDs of the mean, and 99.7% of all observations lie within 3 SDs of the mean.

For competitive gymnasts in this sample, the **mean age at menarche is 13.4** years, with a **SD of 1.3** years. Based on the 68-95-99.7 rule:

- 68% of observations lie within 1 SD: 13.4 ± 1.3 = 12.1-14.7.
- **95%** of observations lie within **2 SDs**: 13.4 ± 2.6 = **10.8-16.0**.
- 99.7% of observations lie within 3 SDs: 13.4 ± 3.9 = 9.5-17.3.

An onset of menarche at **age ≥16 years is 2 SDs** from the mean; therefore, **2.5%** of the observations must lie **above 16 years** (with 2.5% of observations below 10.8 years). The probability that a random competitive gymnast will have an onset of menarche at age ≥16 years is 0.025.

(Choices A, B, and C) The 68-95-99.7 rule states that 99.7% of observations lie within 3 SDs; therefore, 0.997 is the probability that a random competitive gymnast will have an onset of menarche between age 9.5 and 17.3 years. Similarly, 0.95 is the probability that a random competitive gymnast will have an onset of menarche between age 10.8 and 16.0 years (ie, within ± 2 SDs from the mean), and 0.68 is the probability that a random competitive gymnast will have an onset of menarche between age 12.1 and 14.7 years (ie, within ± 1 SD from the mean).

(Choices D and F) Based on the 68-95-99.7 rule, 32% (ie, 100% − 68%) of observations lie outside 1 SD from the mean, with half (ie, 32/2 = 16%) above and half (16%) below 1 SD from the mean. Therefore, 0.160 is the probability that a random competitive gymnast will have an onset of menarche at either age ≤12.1 years (ie, ≥1 SD below the mean) or at age ≥14.7 years (ie, ≥1 SD above the mean). Similarly, 0.0015 is the probability that a random competitive gymnast will have an onset of menarche at either age ≤9.5 years (ie, ≥3 SDs below the mean) or at age ≥17.3 years (ie, ≥3 SDs above the mean).

Educational objective:
In a normal (bell-shaped) distribution, 68% of all values are within 1 standard deviation (SD) of the mean; 95% are within 2 SDs of the mean; and 99.7% are within 3 SDs of the mean.

Biostatistics Biostatistics & Epidemiology Normal distribution

99.7% of all observations lie within 3 SDs of the mean.

For competitive gymnasts in this sample, the **mean age at menarche is 13.4** years, with a **SD of 1.3** years. Based on the 68-95-99.7 rule:

- 68% of observations lie within 1 SD: 13.4 ± 1.3 = 12.1-14.7.
- **95%** of observations lie within **2 SDs**: 13.4 ± 2.6 = **10.8-16.0**.
- 99.7% of observations lie within 3 SDs: 13.4 ± 3.9 = 9.5-17.3.

An onset of menarche at **age ≥16 years is 2 SDs** from the mean; therefore, **2.5%** of the observations must lie **above 16 years** (with 2.5% of observations below 10.8 years). The probability that a random competitive gymnast will have an onset of menarche at age ≥16 years is 0.025.

(Choices A, B, and C) The 68-95-99.7 rule states that 99.7% of observations lie within 3 SDs; therefore, 0.997 is the probability that a random competitive gymnast will have an onset of menarche between age 9.5 and 17.3 years. Similarly, 0.95 is the probability that a random competitive gymnast will have an onset of menarche between age 10.8 and 16.0 years (ie, within ± 2 SDs from the mean), and 0.68 is the probability that a random competitive gymnast will have an onset of menarche between age 12.1 and 14.7 years (ie, within ± 1 SD from the mean).

(Choices D and F) Based on the 68-95-99.7 rule, 32% (ie, 100% − 68%) of observations lie outside 1 SD from the mean, with half (ie, 32/2 = 16%) above and half (16%) below 1 SD from the mean. Therefore, 0.160 is the probability that a random competitive gymnast will have an onset of menarche at either age ≤12.1 years (ie, ≥1 SD below the mean) or at age ≥14.7 years (ie, ≥1 SD above the mean). Similarly, 0.0015 is the probability that a random competitive gymnast will have an onset of menarche at either age ≤9.5 years (ie, ≥3 SDs below the mean) or at age ≥17.3 years (ie, ≥3 SDs above the mean).

Educational objective:
In a normal (bell-shaped) distribution, 68% of all values are within 1 standard deviation (SD) of the mean; 95% are within 2 SDs of the mean; and 99.7% are within 3 SDs of the mean.

Biostatistics	Biostatistics & Epidemiology	Normal distribution
Subject	System	Topic

Copyright © UWorld. All rights reserved.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Researchers are interested in the association between colorectal carcinoma and nonsteroidal anti-inflammatory drug use. They first interview a group of patients with biopsy-proven colorectal carcinoma and then interview a group consisting of the patients' neighbors who are of similar age and race. The analysis is based on comparing the results of pairs of individuals (one from each of the 2 groups) who have similar characteristics. This design technique best helps address which of the following potential problems with this study?

A. Ascertainment bias

B. Confounding

C. Observer bias

D. Recall bias

E. Selection bias

Submit

Researchers are interested in the association between colorectal carcinoma and nonsteroidal anti-inflammatory drug use. They first interview a group of patients with biopsy-proven colorectal carcinoma and then interview a group consisting of the patients' neighbors who are of similar age and race. The analysis is based on comparing the results of pairs of individuals (one from each of the 2 groups) who have similar characteristics. This design technique best helps address which of the following potential problems with this study?

- A. Ascertainment bias (3%)
- ✓ B. Confounding (45%)
- C. Observer bias (2%)
- D. Recall bias (5%)
- E. Selection bias (43%)

Omitted
Correct answer
B

45%
Answered correctly

02 secs
Time Spent

02/18/2020
Last Updated

Explanation

Matching is a method generally used in the design stage of case-control studies to control confounding (ie, when a perceived association between an exposure and an outcome is actually explained by a confounding variable associated with both the exposure and the outcome). The initial step in matching involves selecting variables that could be confounders (eg, age, race). Cases and controls are then selected based on the matching variables so that both groups have a **similar distribution** in accordance with the variables.

In this scenario, the "cases" (patients with colorectal cancer) were matched with neighborhood "controls" of similar age and race. Selecting neighbors as controls has another advantage of matching the cases to controls by variables that are difficult to measure (eg, socioeconomic

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Explanation

Matching is a method generally used in the design stage of case-control studies to control confounding (ie, when a perceived association between an exposure and an outcome is actually explained by a confounding variable associated with both the exposure and the outcome). The initial step in matching involves selecting variables that could be confounders (eg, age, race). Cases and controls are then selected based on the matching variables so that both groups have a **similar distribution** in accordance with the variables.

In this scenario, the "cases" (patients with colorectal cancer) were matched with neighborhood "controls" of similar age and race. Selecting neighbors as controls has another advantage of matching the cases to controls by variables that are difficult to measure (eg, socioeconomic status, environmental factors). Gender and smoking status are other common confounders.

(Choices A, C, D, and E) Observer bias and ascertainment bias result from mislabeling exposed/unexposed or cases/controls. Recall bias could be a limitation of this study as the interviewed participants with colorectal cancer may be more likely to recall certain exposures. Selection bias is a potential problem in this study because the controls selected may not reflect the exposure experience of the general population. However, although these biases may be present, matching best addresses confounding rather than any of these biases.

Educational objective:
Matching is used in case-control studies in order to control confounding. Matching variables should always be the potential confounders of the study (eg, age, race). Cases and controls are then selected based on the matching variables so that both groups have a similar distribution in accordance with the variables.

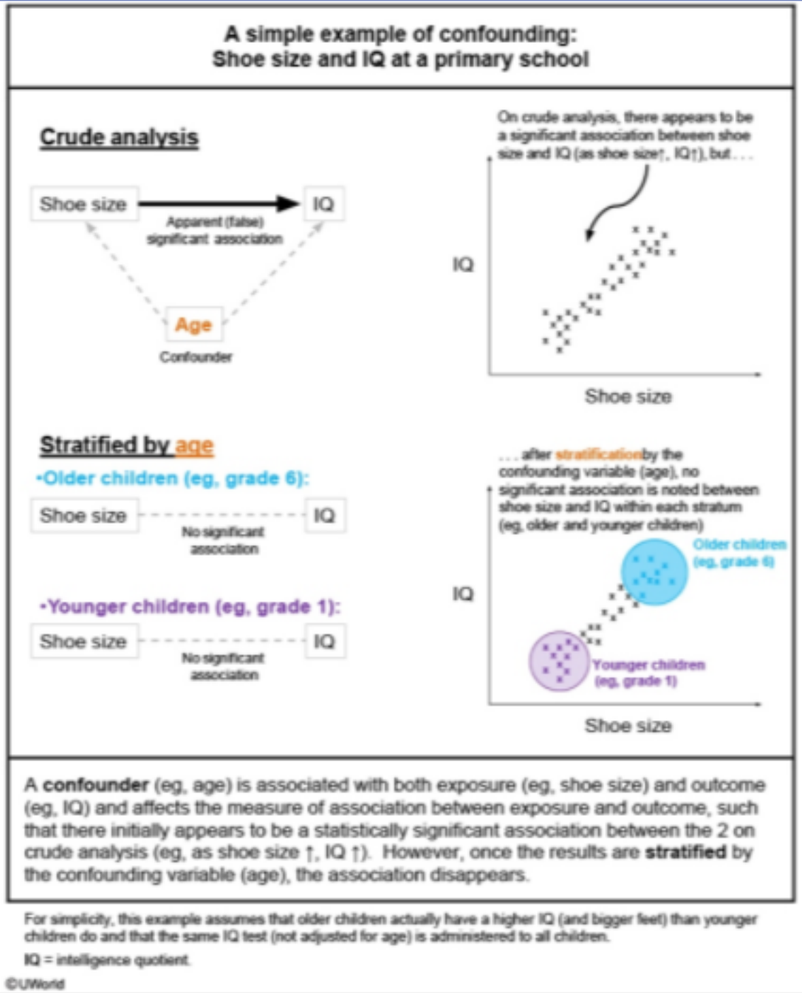
References

- Matching.

Biostatistics	Biostatistics & Epidemiology	Bias
Subject	System	Topic

Copyright © UWorld. All rights reserved.

Exhibit Display



Zoom In | Zoom Out | Reset | Add To New Card | Existing Card

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

A study aims to investigate the effectiveness of the topical application of tetracycline ointment in the prevention of surgical-site infection associated with resection of advanced oral cavity cancer. Fifty patients who underwent locoregional resection were assigned by chance to receive standard care alone or tetracycline-ointment in addition to standard care. The patients were then followed for 1 month and the number and severity of surgical-site infections were recorded for each group. Which of the following best describes the study design?

A. Case-control study

B. Crossover design

C. Prospective cohort study

D. Randomized control trial

E. Retrospective cohort study

Submit

A study aims to investigate the effectiveness of the topical application of tetracycline ointment in the prevention of surgical-site infection associated with resection of advanced oral cavity cancer. Fifty patients who underwent locoregional resection were assigned by chance to receive standard care alone or tetracycline-ointment in addition to standard care. The patients were then followed for 1 month and the number and severity of surgical-site infections were recorded for each group. Which of the following best describes the study design?

- A. Case-control study (7%)
- B. Crossover design (0%)
- C. Prospective cohort study (23%)
- ✓ D. Randomized control trial (68%)
- E. Retrospective cohort study (0%)

Omitted
Correct answer
D

68%
Answered correctly

03 secs
Time Spent

05/25/2020
Last Updated

Explanation

Research studies are classified as observational studies (eg, cross-sectional, case control, cohort) or experimental studies (eg, randomized control trials, factorial, crossover), depending on how much control a researcher has over the independent variables (eg, risk factors/exposures, interventions) in a study. The main differences between these 2 types of study designs are as follows:

- In **observational designs**, the researcher **observes the effect** of naturally occurring risk factors/exposures on outcomes of interest.
- In **experimental designs**, the researcher **randomly assigns interventions** to potential participants to assess the effect of the controlled interventions

Research studies are classified as observational studies (eg, cross-sectional, case control, cohort) or experimental studies (eg, randomized control trials, factorial, crossover), depending on how much control a researcher has over the independent variables (eg, risk factors/exposures, interventions) in a study. The main differences between these 2 types of study designs are as follows:

- In **observational designs**, the researcher **observes the effect** of naturally occurring risk factors/exposures on outcomes of interest.
- In **experimental designs**, the researcher **randomly assigns interventions** to potential participants to assess the effect of the controlled interventions.

A **randomized control trial** is one of the simplest subtypes of experimental designs; it is an experiment in which **participants** are **randomly allocated** to **≥2 groups** to assess the effect of **specific interventions** (eg, treatments). In this example, 50 patients who underwent surgical resection of advanced oral cavity cancer were assigned by chance into either a tetracycline-ointment treatment group or a standard-of-care control group. The effectiveness of the topical application of tetracycline ointment for prevention of surgical-site infection was then assessed by recording the incidence and severity of infections in each group.

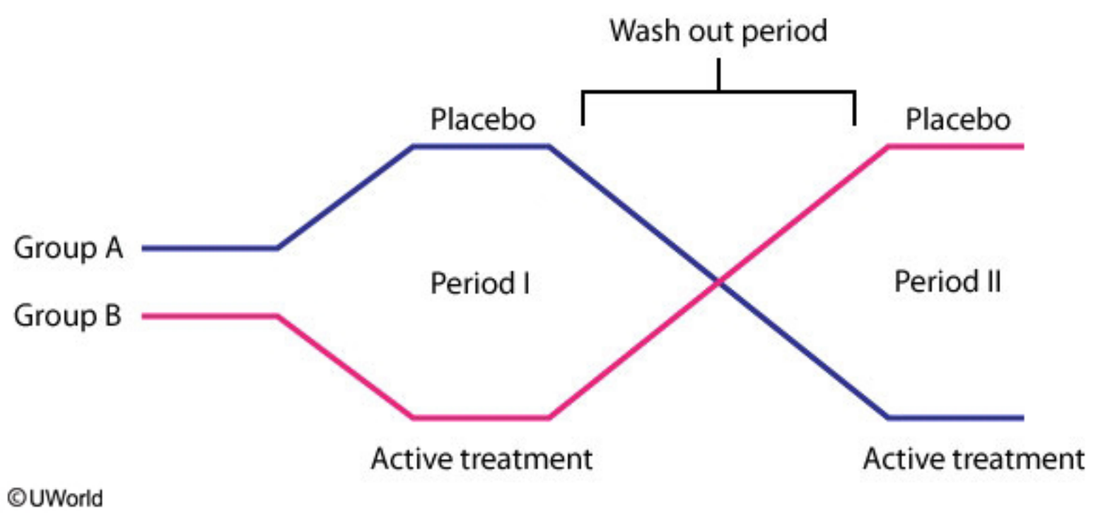
(Choice A) A case-control study is a type of observational study in which potential participants are initially identified as cases or controls according to outcome status (ie, developing the disease or condition of interest); the 2 groups are then assessed on their past exposure to specific risk factors.

(Choice B) A crossover study is an experiment in which subjects are exposed to different treatments or exposures sequentially. In the above scenario, patients are exposed to a single treatment only.

(Choices C and E) In prospective cohort studies, exposure status is determined in the present and patients are tracked over time for development of the outcome of interest; in retrospective cohort studies, exposure status is determined at a particular point in the past, and participants are tracked retrospectively, typically through their medical records. However, participants in cohort studies already have a definite exposure status; they are not randomly assigned to exposure groups as in the above example.

Educational objective:
A randomized control trial is an experiment in which participants are randomly allocated to ≥2 groups to assess the effect of specific interventions (eg, treatments).

Exhibit Display



Zoom In
Zoom Out
Reset
Add To Flash Card

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

A prospective cohort study was conducted to assess the role of daily alcohol consumption in the occurrence of breast carcinoma. The investigators reported a 5-year relative risk of 1.4 for people who consume alcohol daily compared to those who do not. The 95% confidence interval was 1.02-1.85. Which of the following p-values is most consistent with the results described above?

A. 0.03
B. 0.06
C. 0.09
D. 0.11
E. 0.20

Submit

A prospective cohort study was conducted to assess the role of daily alcohol consumption in the occurrence of breast carcinoma. The investigators reported a 5-year relative risk of 1.4 for people who consume alcohol daily compared to those who do not. The 95% confidence interval was 1.02-1.85. Which of the following p-values is most consistent with the results described above?

- A. 0.03 (68%)
- B. 0.06 (15%)
- C. 0.09 (6%)
- D. 0.11 (3%)
- E. 0.20 (5%)

Omitted
Correct answer
A

68%
Answered correctly

02 secs
Time Spent

01/13/2020
Last Updated

Explanation

Relative risk (RR) is used in cohort studies to determine how strongly a risk factor (ie, exposure) is associated with an outcome. RR is the risk of an outcome (eg, breast cancer) in the exposed group (eg, individuals who consume alcohol daily) divided by the risk of that outcome in the unexposed group (eg, individuals who do not consume alcohol daily). If the RR = 1.0 (null value), then there is no association between the exposure and the disease. An RR >1.0 indicates that the exposure is associated with increased risk of disease. An RR <1.0 means that the exposure is associated with decreased risk of disease.

The RR by itself does not account for the possibility that chance alone is responsible for the results. The 95% confidence interval (CI) and p-value are 2 measures of **statistical significance** that can help strengthen the findings of a study using RR. For a result to be considered statistically

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Explanation

Relative risk (RR) is used in cohort studies to determine how strongly a risk factor (ie, exposure) is associated with an outcome. RR is the risk of an outcome (eg, breast cancer) in the exposed group (eg, individuals who consume alcohol daily) divided by the risk of that outcome in the unexposed group (eg, individuals who do not consume alcohol daily). If the RR = 1.0 (null value), then there is no association between the exposure and the disease. An RR >1.0 indicates that the exposure is associated with increased risk of disease. An RR <1.0 means that the exposure is associated with decreased risk of disease.

The RR by itself does not account for the possibility that chance alone is responsible for the results. The 95% confidence interval (CI) and p-value are 2 measures of **statistical significance** that can help strengthen the findings of a study using RR. For a result to be considered statistically significant, its corresponding CI must NOT contain the null value. When the **95% CI** does **not** include the null value, this gives a corresponding **p-value <0.05** and the association between exposure and outcome is considered statistically significant. A p-value <0.05 reflects that there is a very low probability that the result was due to chance alone; formally, the p-value is the probability of observing a given (or more extreme) result due to chance alone assuming that the null hypothesis is true.

In this example, the RR is 1.4 with a 95% CI of 1.02-1.85. It can be concluded that daily alcohol consumption is associated with an increased risk of breast carcinoma (RR >1) and that the findings are statistically significant (95% CI does not include the null value of 1.0). Therefore, the expected p-value would be <0.05.

(Choices B, C, D, and E) These options contain p-values >0.05, so the results would not be statistically significant. Note the relationship between CI and p-value: a statistically significant 95% CI corresponds to a p-value <0.05; a statistically significant 99% CI (would also not include the null value and likely be wider than a 95% CI) corresponds to a p-value <0.01.

Educational objective:
A result is considered statistically significant if the 95% confidence interval does not cross the null value, which corresponds to a p-value <0.05.

Biostatistics	Biostatistics & Epidemiology	P-value and confidence interval
Subject	System	Topic

Copyright © UWorld. All rights reserved.

Explanation

Exhibit Display

Statistical significance: Relationship between confidence intervals and p-values	
A ____ confidence interval that does not cross the null value ...	... corresponds to a p-value <____
95%	0.05
99%	0.01

©UWorld

Zoom In | Zoom Out | Reset | Add To New Card | Existing Card

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

A research group conducted a placebo-controlled clinical trial to assess whether a new drug to treat acute migraine with or without aura in adults is more effective than standard therapy. A total of 3,500 patients with acute migraine were enrolled in the study and randomly assigned to either the new drug or standard treatment. During the data analysis phase, the researchers decide to set alpha at 0.01 rather than 0.05. Which of the following is the most likely result of this change?

A. Any significant findings will be reported with greater confidence

B. There will be a higher probability of a type I error

C. There will be a higher probability of finding statistically significant results

D. There will be a lower probability of a type II error

E. The study will have more statistical power

Submit

A research group conducted a placebo-controlled clinical trial to assess whether a new drug to treat acute migraine with or without aura in adults is more effective than standard therapy. A total of 3,500 patients with acute migraine were enrolled in the study and randomly assigned to either the new drug or standard treatment. During the data analysis phase, the researchers decide to set alpha at 0.01 rather than 0.05. Which of the following is the most likely result of this change?

- ✓ A. Any significant findings will be reported with greater confidence (64%)
- B. There will be a higher probability of a type I error (10%)
- C. There will be a higher probability of finding statistically significant results (10%)
- D. There will be a lower probability of a type II error (5%)
- E. The study will have more statistical power (8%)

Omitted
Correct answer
A

64%
Answered correctly

02 secs
Time Spent

05/12/2020
Last Updated

Explanation

Type I (α) and type II (β) errors

	True status	
	There is a true difference	There is NO true difference

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

Type I (α) and type II (β) errors

		True status	
		There is a true difference (ie, H_0 is false)	**There is NO true difference** (ie, H_0 is true)
Study result	**Difference calculated as statistically significant** (ie, reject H_0)	**Correctly** conclude there is a difference	**Type I (α) error** (Falsely conclude there is a difference)
	Difference calculated as NOT statistically significant (ie, fail to reject H_0)	**Type II (β) error** (Falsely conclude there is NO difference)	**Correctly** conclude there is NO difference

H_0 = null hypothesis of no difference.

©UWorld

Alpha (*α*) is the significance level used to establish the **statistical significance** of study results. It is usually predetermined by investigators as the threshold to reject a null hypothesis (H_0). The results of a study are considered statistically significant when the *p*-value (ie, probability of obtaining results as extreme as the observed results when H_0 is true) is less than the significance level (ie, ***p*-value < α**). Alpha is also the probability of making a **type I error**, which is rejecting a true H_0 (ie, a false positive: finding a statistically significant difference when one does not truly exist). The **complement** of alpha/significance level is the **confidence level** ($1 - \alpha$), which represents the probability of not rejecting a true H_0 (ie, a true negative, or not finding a statistically significant difference when one does not truly exist).

H_0 = null hypothesis of no difference.

©UWorld

Alpha (α) is the significance level used to establish the **statistical significance** of study results. It is usually predetermined by investigators as the threshold to reject a null hypothesis (H_0). The results of a study are considered statistically significant when the *p*-value (ie, probability of obtaining results as extreme as the observed results when H_0 is true) is less than the significance level (ie, ***p*-value < α**). Alpha is also the probability of making a **type I error**, which is rejecting a true H_0 (ie, a false positive: finding a statistically significant difference when one does not truly exist). The **complement** of alpha/significance level is the **confidence level** ($1 - \alpha$), which represents the probability of not rejecting a true H_0 (ie, a true negative, or not finding a statistically significant difference when one does not truly exist).

The lower the significance level, the lower the probability of a type I error. However, this means that the threshold to attain statistical significance is more stringent. Therefore, **reducing the significance level** (eg, from $\alpha = 5\%$ to $\alpha = 1\%$) **decreases** the probability of a **type I error** along with the probability of **finding statistical significance** (eg, *p*-value < 0.05 to *p*-value < 0.01) **(Choices B and C)**. A decreased significance level also means an **increased confidence level** ($1 - \alpha$); therefore, any statistically significant findings can be reported with a greater level of confidence.

(Choices D and E) A type II error occurs when a false H_0 is not rejected (ie, a false negative: not finding a statistically significant difference when one truly exist). The probability of a type II error (known as beta [β]) is the complement of power ($1 - \beta$). Statistical power is the probability of rejecting a false H_0. That is, the power to find a statistically significant difference when one truly exists (ie, true positive). Reducing the significance level from 0.05 to 0.01 makes the threshold to attain statistical significance more stringent. Consequently, the study will have less statistical power to detect a statistically significant difference when one truly exists and will have a higher probability of a type II error (the complement of power).

Educational objective:
Reducing the significance level alpha (α) in a study allows researchers to report any significant findings with greater confidence.

Biostatistics	Biostatistics & Epidemiology	Hypothesis testing
Subject	System	Topic

Copyright © UWorld. All rights reserved.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

A new portable cholesterol-measuring device is being developed for use in medically underserved communities. During clinical trials of an early prototype, a patient's cholesterol level is found to be 200 mg/dL, 201 mg/dL, and 200 mg/dL on 3 separate measurements of the same blood sample. Using the gold standard measurement method, the same sample is found to have a cholesterol level of 260 mg/dL. Which of the following descriptions best characterizes the new cholesterol-measuring device?

A. High accuracy; high precision
B. High accuracy; low precision
C. High sensitivity; low specificity
D. Low accuracy; high precision
E. Low accuracy; low precision
F. Low sensitivity; high specificity

Submit

A new portable cholesterol-measuring device is being developed for use in medically underserved communities. During clinical trials of an early prototype, a patient's cholesterol level is found to be 200 mg/dL, 201 mg/dL, and 200 mg/dL on 3 separate measurements of the same blood sample. Using the gold standard measurement method, the same sample is found to have a cholesterol level of 260 mg/dL. Which of the following descriptions best characterizes the new cholesterol-measuring device?

- A. High accuracy; high precision (0%)
- B. High accuracy; low precision (7%)
- C. High sensitivity; low specificity (0%)
- ✓ D. Low accuracy; high precision (89%)
- E. Low accuracy; low precision (1%)
- F. Low sensitivity; high specificity (0%)

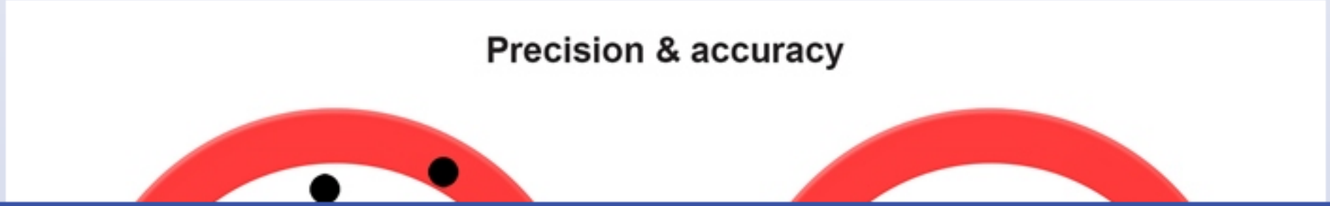
Omitted
Correct answer
D

89%
Answered correctly

03 secs
Time Spent

05/08/2020
Last Updated

Explanation



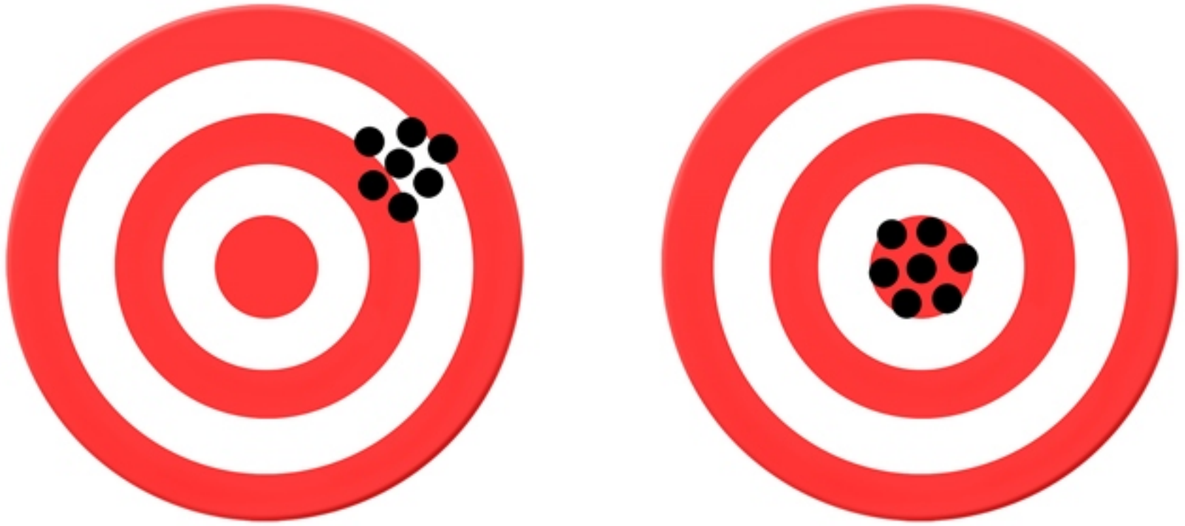
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

Precision & accuracy

Low precision
Low accuracy

Low precision
High accuracy

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31



©UWorld

- **Precision** (reliability) is the ability of a test to reproduce identical or similar results with repeated measurements (bottom half of figure).
- **Accuracy** (validity) is the ability of a test to measure what it is supposed to measure. For a new test to be accurate, its results should be equivalent to the results obtained with a gold standard (eg, best conventional test available) on the same individual (right half of figure).

In this case, repeated measurements of the same blood sample yielded **nearly identical values**; therefore, the new device has **high precision**. However, none of the measurements with the new device were equivalent to the results obtained with the gold standard (ie, the measurements are **incorrect**), meaning the new device has **low accuracy**.

(Choices C and F) Sensitivity is defined as the percentage of diseased individuals who are correctly identified (ie, proportion of subjects who have a positive test among all those with disease), whereas specificity is defined as the percentage of healthy individuals who are correctly identified (ie, proportion of subjects who have a negative test among all those without disease). Poor accuracy or precision can lead to low test sensitivity and specificity. However, no information is provided to determine the sensitivity or specificity of the new device compared to the gold

- **Precision** (reliability) is the ability of a test to reproduce identical or similar results with repeated measurements (bottom half of figure).
- **Accuracy** (validity) is the ability of a test to measure what it is supposed to measure. For a new test to be accurate, its results should be equivalent to the results obtained with a gold standard (eg, best conventional test available) on the same individual (right half of figure).

In this case, repeated measurements of the same blood sample yielded **nearly identical values**; therefore, the new device has **high precision**. However, none of the measurements with the new device were equivalent to the results obtained with the gold standard (ie, the measurements are **incorrect**), meaning the new device has **low accuracy**.

(Choices C and F) Sensitivity is defined as the percentage of diseased individuals who are correctly identified (ie, proportion of subjects who have a positive test among all those with disease), whereas specificity is defined as the percentage of healthy individuals who are correctly identified (ie, proportion of subjects who have a negative test among all those without disease). Poor accuracy or precision can lead to low test sensitivity and specificity. However, no information is provided to determine the sensitivity or specificity of the new device compared to the gold standard test.

Educational objective:
A precise tool is one that consistently provides very similar or the same value when measuring a fixed quantity. An accurate tool is one that provides a measurement identical or similar to the actual value (as reflected in a gold standard measurement).

Biostatistics	Biostatistics & Epidemiology	Accuracy and precision
Subject	System	Topic

Copyright © UWorld. All rights reserved.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Researchers want to study the effects of labyrinthectomy with cochlear implantation on hearing, vertigo, and tinnitus. A random sample of medical charts is selected from a cohort of patients who had undergone labyrinthectomy with cochlear implantation in the same ear for intractable vertigo and hearing loss and who had documentation of both preoperative and postoperative audiometric evaluations. Tinnitus is quantified using the Tinnitus Handicap Inventory (THI) before and after the interventions. The THI score ranges from 0 to 100, with higher values indicating greater tinnitus severity. Which of the following statistical tests is most appropriate for comparing preoperative and postoperative THI scores?

A. Analysis of variance

B. Chi-square test

C. Correlation analysis

D. Meta-analysis

E. Paired *t*-test

Submit

Researchers want to study the effects of labyrinthectomy with cochlear implantation on hearing, vertigo, and tinnitus. A random sample of medical charts is selected from a cohort of patients who had undergone labyrinthectomy with cochlear implantation in the same ear for intractable vertigo and hearing loss and who had documentation of both preoperative and postoperative audiometric evaluations. Tinnitus is quantified using the Tinnitus Handicap Inventory (THI) before and after the interventions. The THI score ranges from 0 to 100, with higher values indicating greater tinnitus severity. Which of the following statistical tests is most appropriate for comparing preoperative and postoperative THI scores?

- A. Analysis of variance (4%)
- B. Chi-square test (11%)
- C. Correlation analysis (13%)
- D. Meta-analysis (1%)
- ✓ E. Paired *t*-test (69%)

Omitted
Correct answer
E

69%
Answered correctly

01 sec
Time Spent

05/30/2020
Last Updated

Explanation

		Dependent variable	
		Qualitative (categorical)	*Quantitative*
Independent variable	*Qualitative (categorical)*	Chi-square, logistic regression*	*t*-test, ANOVA, linear regression

		Dependent variable	
		Qualitative (categorical)	*Quantitative*
Independent variable	*Qualitative (categorical)*	Chi-square, logistic regression*	*t*-test, ANOVA, linear regression
	Quantitative	Logistic regression*	Correlation, linear regression

* Dependent variable must be dichotomous.

ANOVA = analysis of variance.

Variables are broadly classified as qualitative (ie, categorical) or quantitative (ie, continuous) based on their scale of measurement. **Qualitative variables** (eg, type of treatment, blood type) represent categories or groups, whereas **quantitative variables** (eg, temperature, glucose levels) represent numerical values. The scale of measurement of the dependent (eg, outcome) and independent (eg, exposures, risk factors) variables in a study determines the correct statistical test for any given situation.

A ***t*-test** compares the **mean of 2 groups**. It requires that a quantitative dependent variable (ie, outcome) be evaluated in 2 groups that are classified based upon a categorical independent variable (ie, exposure).

In this study:

- The quantitative dependent variable was the **Tinnitus Handicap Inventory (THI) score**.
- The categorical independent variable was the **audiometric evaluation** (with categories "preoperative" and "postoperative").

A *t*-test will determine whether there is a statistically significant difference in mean THI scores between the preoperative and postoperative audiometric evaluations. A large, statistically significant difference in mean scores indicates that labyrinthectomy with cochlear implantation is associated with changes in tinnitus (ie, the null hypothesis is rejected).

There are 2 types of *t*-tests: the independent samples *t*-test (used when 2 groups are independent) and the **paired *t*-test** (used with 2 related groups with matched pairs). Matched groups are formed when each observation in one group is paired with an observation from the other group. Examples include a study in which individuals have been assessed twice (eg, before and after an intervention) and one in which 2 groups of individuals have been matched based on certain attributes (eg, age, severity of disease). In this case, THI scores are assessed twice for each

- The categorical independent variable was the **audiometric evaluation** (with categories "preoperative" and "postoperative").

A *t*-test will determine whether there is a statistically significant difference in mean THI scores between the preoperative and postoperative audiometric evaluations. A large, statistically significant difference in mean scores indicates that labyrinthectomy with cochlear implantation is associated with changes in tinnitus (ie, the null hypothesis is rejected).

There are 2 types of *t*-tests: the independent samples *t*-test (used when 2 groups are independent) and the **paired *t*-test** (used with 2 related groups with matched pairs). Matched groups are formed when each observation in one group is paired with an observation from the other group. Examples include a study in which individuals have been assessed twice (eg, before and after an intervention) and one in which 2 groups of individuals have been matched based on certain attributes (eg, age, severity of disease). In this case, THI scores are assessed twice for each patient (preoperative and postoperative audiometric evaluations); therefore, a paired *t*-test is used.

(Choice A) The analysis of variance (ANOVA) test compares the mean of ≥3 independent groups, as in a study comparing serum ferritin concentrations (ie, quantitative variable) in children (age 0-17), adults (age 18-59), and seniors (age ≥60).

(Choice B) The chi-square test evaluates the association between 2 categorical variables, as in a study evaluating the association between sex (ie, "male" and "female") and myocardial infarction (ie, presence or absence).

(Choice C) A correlation analysis uses the correlation coefficient to describe the linear relationship between 2 quantitative variables, as in a study evaluating the linear relationship between hours of sleep and irritability score.

(Choice D) Meta-analysis is a statistical technique used to combine and analyze data from several studies to conduct an analysis with a greater statistical power than that of the individual studies.

Educational objective:
The paired *t*-test compares the mean of 2 related groups. The test requires that a quantitative dependent variable (ie, outcome) be evaluated in 2 related (ie, matched, paired) groups.

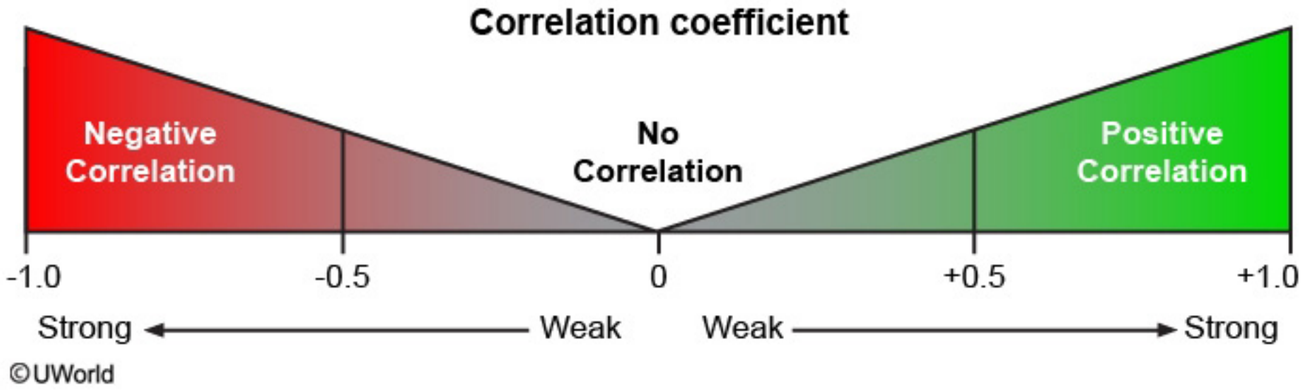
References

- The differences and similarities between two-sample t-test and paired t-test.

Biostatistics | Biostatistics & Epidemiology | Statistical tests

- The categorical independent variable was the **audiometric evaluation** (with categories "preoperative" and "postoperative").

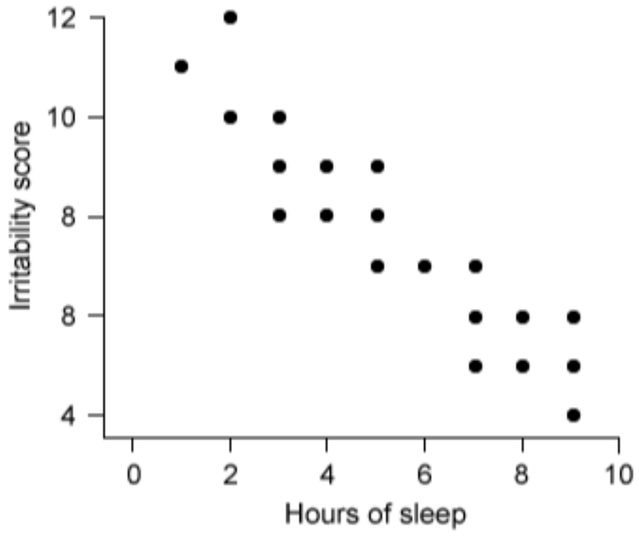
Exhibit Display



Zoom In
Zoom Out
Reset
Add To Flash Card

- The categorical independent variable was the **audiometric evaluation** (with categories "preoperative" and "postoperative").

Exhibit Display



Zoom In Zoom Out Reset Add To Flash Card

A study is conducted to assess whether quantitative measurement of heat shock protein 90 alpha (Hsp90α) in plasma can improve diagnostic accuracy and help monitor treatment response in liver cancer patients. A total of 300 individuals (100 liver cancer patients and 200 healthy controls) participate in the study. Analyses reveal that 90 of the liver cancer patients have an abnormally high plasma Hsp90α level and that 12 of the healthy controls have an abnormally high plasma Hsp90α level. An abnormally high plasma Hsp90α level indicates a positive test for liver cancer. Which of the following values best represents the specificity of this test?

- A. 6%
- B. 10%
- C. 33%
- D. 90%
- E. 94%

Submit

A study is conducted to assess whether quantitative measurement of heat shock protein 90 alpha (Hsp90α) in plasma can improve diagnostic accuracy and help monitor treatment response in liver cancer patients. A total of 300 individuals (100 liver cancer patients and 200 healthy controls) participate in the study. Analyses reveal that 90 of the liver cancer patients have an abnormally high plasma Hsp90α level and that 12 of the healthy controls have an abnormally high plasma Hsp90α level. An abnormally high plasma Hsp90α level indicates a positive test for liver cancer. Which of the following values best represents the specificity of this test?

- A. 6% (5%)
- B. 10% (1%)
- C. 33% (2%)
- D. 90% (9%)
- ✓ E. 94% (81%)

Omitted
Correct answer
E

81%
Answered correctly

02 secs
Time Spent

05/28/2020
Last Updated

Explanation

	Positive condition	Negative condition	
Positive test result	TP	FP	PPV = TP/(TP + FP)
			NPV =

	Positive condition	Negative condition	
Positive test result	TP	FP	PPV = TP/(TP + FP)
Negative test result	FN	TN	NPV = TN/(TN + FN)
	Sensitivity = TP/(TP + FN)	Specificity = TN/(TN + FP)	

FN = false negative; **FP** = false positive; **NPV** = negative predictive value; **PPV** = positive predictive value; **TN** = true negative; **TP** = true positive.

The **specificity** of a test refers to its ability to correctly identify individuals without the disease. Specificity is equal to the number of individuals who do not have the disease and who test negative (true negatives [TN]) divided by the total number of individuals without the disease (TN + false positives [FP]):

Specificity = TN / (TN + FP)

In this example, 12 out of 200 healthy controls had an abnormally high plasma heat shock protein 90 alpha (Hsp90α) level (ie, tested positive); therefore, FP = 12 and TN = 188 (ie, 200 − 12). Therefore, the specificity of the test is as follows:

Specificity = TN / (TN + FP) = 188 / (188 + 12) = 188 / 200 = 0.94 (or **94%**)

(Choice A) In this example, 6% is the false positive rate (FPR): FPR = FP / (TN + FP) = 12 / (188 + 12) = 12 / 200 = 0.06 (or 6%). FPR is the probability of a positive test in a patient without the disease (ie, FP); it is also the complement to specificity (ie, FPR = 1 − specificity = 1 − 0.94 = 0.06).

(Choices B and D) The sensitivity of a test refers to its ability to correctly identify individuals with the disease. It is calculated as the number of individuals who have the disease and who test positive (true positives [TP]) divided by the total number of individuals with the disease (TP + false negatives [FN]): sensitivity = TP / (TP + FN). In this example, 90 of 100 liver cancer patients had an abnormally high plasma Hsp90α level (ie, tested positive); therefore, TP = 90 and FN = 10 (ie, 100 − 90). Accordingly, the sensitivity of the test is 90 / (90 + 10) = 90 / 100 = 0.90 (or 90%).

positives [FP]):

Specificity = TN / (TN + FP)

In this example, 12 out of 200 healthy controls had an abnormally high plasma heat shock protein 90 alpha (Hsp90α) level (ie, tested positive); therefore, FP = 12 and TN = 188 (ie, 200 − 12). Therefore, the specificity of the test is as follows:

Specificity = TN / (TN + FP) = 188 / (188 + 12) = 188 / 200 = 0.94 (or **94%**)

(Choice A) In this example, 6% is the false positive rate (FPR): FPR = FP / (TN + FP) = 12 / (188 + 12) = 12 / 200 = 0.06 (or 6%). FPR is the probability of a positive test in a patient without the disease (ie, FP); it is also the complement to specificity (ie, FPR = 1 − specificity = 1 − 0.94 = 0.06).

(Choices B and D) The sensitivity of a test refers to its ability to correctly identify individuals with the disease. It is calculated as the number of individuals who have the disease and who test positive (true positives [TP]) divided by the total number of individuals with the disease (TP + false negatives [FN]): sensitivity = TP / (TP + FN). In this example, 90 of 100 liver cancer patients had an abnormally high plasma Hsp90α level (ie, tested positive); therefore, TP = 90 and FN = 10 (ie, 100 − 90). Accordingly, the sensitivity of the test is 90 / (90 + 10) = 90 / 100 = 0.90 (or 90%). Sensitivity is related to the false negative rate (FNR): FNR = 1 – sensitivity. FNR is the probability of a negative test result in a patient with the disease (ie, FN): FNR = FN / (TP + FN) = 10 / (10 + 90) = 10 / 100 = 0.10 (or 10%).

(Choice C) In this example, 33% represents the proportion of study participants with liver cancer: 100 liver cancer patients / (100 liver cancer patients + 200 health controls) = 0.333 (or 33.3%).

Educational objective:
The specificity of a test is its ability to correctly identify individuals without the disease. Specificity can be calculated as follows:

Specificity = True negatives / (True negatives + False positives)

Biostatistics	Biostatistics & Epidemiology	Sensitivity and specificity
Subject	System	Topic

Copyright © UWorld. All rights reserved.

A test is used to diagnose dementia in a population of 1,000 women age ≥75. The test has a sensitivity of 80% and a specificity of 90%. The test is then used as a diagnostic tool in 2 other populations of women of the same age: population 1 has a prevalence of dementia of 15%, and population 2 has a prevalence of dementia of 30%. Which of the following best describes how the negative predictive values (NPV) and the positive predictive values (PPV) from populations 1 and 2 relate to each other?

A. NPV and PPV do not change as prevalence changes

B. NPV in population 1 < NPV population 2; PPV in population 1 < PPV population 2

C. NPV in population 1 < NPV population 2; PPV in population 1 > PPV population 2

D. NPV in population 1 > NPV population 2; PPV in population 1 < PPV population 2

E. NPV in population 1 > NPV population 2; PPV in population 1 > PPV population 2

Submit

A test is used to diagnose dementia in a population of 1,000 women age ≥75. The test has a sensitivity of 80% and a specificity of 90%. The test is then used as a diagnostic tool in 2 other populations of women of the same age: population 1 has a prevalence of dementia of 15%, and population 2 has a prevalence of dementia of 30%. Which of the following best describes how the negative predictive values (NPV) and the positive predictive values (PPV) from populations 1 and 2 relate to each other?

A. NPV and PPV do not change as prevalence changes (6%)

B. NPV in population 1 < NPV population 2; PPV in population 1 < PPV population 2 (6%)

C. NPV in population 1 < NPV population 2; PPV in population 1 > PPV population 2 (6%)

✓ D. NPV in population 1 > NPV population 2; PPV in population 1 < PPV population 2 (79%)

E. NPV in population 1 > NPV population 2; PPV in population 1 > PPV population 2 (1%)

Omitted
Correct answer
D

79%
Answered correctly

02 secs
Time Spent

06/24/2020
Last Updated

Explanation

Predictive values are performance measures of diagnostic tests that are **dependent on the prevalence** of disease in a population of interest.

- **Negative predictive value** (NPV) is the probability (ie, likelihood) that an individual truly does not have the disease given a negative test result. NPV is equal to the number of individuals who do not have the disease and who test negative (true negatives [TN]) divided by all those with a negative test result (TN + false negatives [FN]): **NPV** = TN / (TN + FN).
- **Positive predictive value** (PPV) is the probability (ie, likelihood) that an individual truly does have the disease given a positive test result. PPV is equal to the number of individuals who have the disease and who test positive (true positives [TP]) divided by all those with a

Explanation

Predictive values are performance measures of diagnostic tests that are **dependent on the prevalence** of disease in a population of interest.

- **Negative predictive value** (NPV) is the probability (ie, likelihood) that an individual truly does not have the disease given a negative test result. NPV is equal to the number of individuals who do not have the disease and who test negative (true negatives [TN]) divided by all those with a negative test result (TN + false negatives [FN]): **NPV** = TN / (TN + FN).
- **Positive predictive value** (PPV) is the probability (ie, likelihood) that an individual truly does have the disease given a positive test result. PPV is equal to the number of individuals who have the disease and who test positive (true positives [TP]) divided by all those with a positive test result (TP + false positives [FP]): **PPV** = TP / (TP + FP).

Disease prevalence affects NPV by changing the number of TN and FN tests. It affects PPV by changing the number of TP and FP tests **(Choice A)**:

- Populations with exceptionally **low disease prevalence** (eg, prevalence close to 0%) have almost all TN (ie, more TN) with almost no TP (ie, fewer TP), and most positive tests would be FP (ie, more FP) with almost no FN (ie, fewer FN). In this scenario, NPV is close to 100%, and PPV is close to 0%. This explains why, with decreasing prevalence, **NPV increases** (more TN, fewer FN) and **PPV decreases** (fewer TP, more FP).
- Populations with exceptionally **high disease prevalence** (eg, prevalence close to 100%) have almost all TP (ie, more TP) with almost no TN (ie, fewer TN), and most negative tests would be FN (ie, more FN) with almost no FP (ie, fewer FP). In this scenario, NPV is close to 0%, and PPV is close to 100%. This explains why, with increasing prevalence, **NPV decreases** (fewer TN, more FN) and **PPV increases** (more TP, fewer FP).

(Choices B, C, and E) In this example, population 1 has a lower prevalence (15%) of dementia in women age ≥75 compared to population 2 (30%). Therefore, NPV in population 1 > NPV population 2, and PPV in population 1 < PPV population 2.

Educational objective:
Predictive values change depending on the prevalence of disease in a study population. As disease prevalence increases, the positive predictive value increases, and the negative predictive value decreases.

Explanation

Predictive values are perf

- **Negative predictiv**
  result. NPV is equa
  those with a negativ
- **Positive predictive**
  PPV is equal to the
  positive test result (

Disease prevalence affect
A):

- Populations with ex
  (ie, fewer TP), and r
  and PPV is close to
  TP, more FP).
- Populations with ex
  TN (ie, fewer TN), a
  0%, and PPV is clos
  (more TP, fewer FP)

(Choices B, C, and E) In
(30%). Therefore, NPV in

Educational objective:
Predictive values change
value increases, and the negative predictive value decreases.

Exhibit Display

Common performance measures for diagnostic tests		
Parameter	**Definition**	**Calculation**
Sensitivity	The probability of a diseased person testing positive	Sensitivity = True positives / (True positives + False negatives)
Specificity	The probability of a nondiseased person testing negative	Specificity = True negatives / (True negatives + False positives)
Positive predictive value	The probability that disease is present given a positive result	PPV = True positives / (True positives + False positives)
Negative predictive value	The probability that disease is absent given a negative result	NPV = True negatives / (True negatives + False negatives)
Positive likelihood ratio	A ratio representing the likelihood of having the disease given a positive result	LR+ = Sensitivity / (1 − Specificity)
Negative likelihood ratio	A ratio representing the likelihood of having the disease given a negative result	LR− = (1 − Sensitivity) / Specificity

Add To Flash Card

- **Positive predictive value** (PPV) is the probability (ie, likelihood) that an individual truly does have the disease given a positive test result. PPV is equal to the number of individuals who have the disease and who test positive (true positives [TP]) divided by all those with a positive test result (TP + false positives [FP]): **PPV** = TP / (TP + FP).

Disease prevalence affects NPV by changing the number of TN and FN tests. It affects PPV by changing the number of TP and FP tests **(Choice A)**:

- Populations with exceptionally **low disease prevalence** (eg, prevalence close to 0%) have almost all TN (ie, more TN) with almost no TP (ie, fewer TP), and most positive tests would be FP (ie, more FP) with almost no FN (ie, fewer FN). In this scenario, NPV is close to 100%, and PPV is close to 0%. This explains why, with decreasing prevalence, **NPV increases** (more TN, fewer FN) and **PPV decreases** (fewer TP, more FP).
- Populations with exceptionally **high disease prevalence** (eg, prevalence close to 100%) have almost all TP (ie, more TP) with almost no TN (ie, fewer TN), and most negative tests would be FN (ie, more FN) with almost no FP (ie, fewer FP). In this scenario, NPV is close to 0%, and PPV is close to 100%. This explains why, with increasing prevalence, **NPV decreases** (fewer TN, more FN) and **PPV increases** (more TP, fewer FP).

(Choices B, C, and E) In this example, population 1 has a lower prevalence (15%) of dementia in women age ≥75 compared to population 2 (30%). Therefore, NPV in population 1 > NPV population 2, and PPV in population 1 < PPV population 2.

Educational objective:
Predictive values change depending on the prevalence of disease in a study population. As disease prevalence increases, the positive predictive value increases, and the negative predictive value decreases.

References

- Diagnostic methods I: sensitivity, specificity, and other measures of accuracy.

Biostatistics	Biostatistics & Epidemiology	Sensitivity, specificity, positive predictive value PPV, negative predictive value NPV
Subject	System	Topic

Copyright © UWorld. All rights reserved.

A hospital wants to estimate the prevalence of diabetic nephropathy in the surrounding population of adults with type 2 diabetes. Kidney biopsy samples are obtained from 500 adult patients with diabetes who receive care at the hospital. The samples are then interpreted by 10 different pathologists, 5 of whom work at the hospital and 5 of whom work at nearby institutions. A preliminary analysis shows that the pathologists who work for the hospital are 3 times more likely to interpret the biopsy samples as diabetic nephropathy compared to those who do not work for the hospital. Which of the following most likely explains this difference in interpretation?

A. Confounding

B. Lead-time bias

C. Observer bias

D. Recall bias

E. Selection bias

Submit

A hospital wants to estimate the prevalence of diabetic nephropathy in the surrounding population of adults with type 2 diabetes. Kidney biopsy samples are obtained from 500 adult patients with diabetes who receive care at the hospital. The samples are then interpreted by 10 different pathologists, 5 of whom work at the hospital and 5 of whom work at nearby institutions. A preliminary analysis shows that the pathologists who work for the hospital are 3 times more likely to interpret the biopsy samples as diabetic nephropathy compared to those who do not work for the hospital. Which of the following most likely explains this difference in interpretation?

- A. Confounding (6%)
- B. Lead-time bias (0%)
- ✓ C. Observer bias (78%)
- D. Recall bias (3%)
- E. Selection bias (11%)

Omitted
Correct answer
C

78%
Answered correctly

02 secs
Time Spent

01/16/2020
Last Updated

Explanation

Observer bias (also known as expectancy bias) occurs when investigators **misclassify data** due to preconceived expectations or **prior knowledge** concerning the study or its participants. This type of bias is particularly important when outcomes are **subjective** (ie, involve personal interpretation of clinical, microscopic, or radiographic findings).

In this case, the pathologists who work at the hospital where the study is being conducted may know that the study is investigating diabetic nephropathy, or they may have access to additional medical records that indicate diabetes status. Conversely, the pathologists at other institutions

Explanation

Observer bias (also known as expectancy bias) occurs when investigators **misclassify data** due to preconceived expectations or **prior knowledge** concerning the study or its participants. This type of bias is particularly important when outcomes are **subjective** (ie, involve personal interpretation of clinical, microscopic, or radiographic findings).

In this case, the pathologists who work at the hospital where the study is being conducted may know that the study is investigating diabetic nephropathy, or they may have access to additional medical records that indicate diabetes status. Conversely, the pathologists at other institutions are more likely to be blinded to the study's objectives and the patients' medical history.

(Choice A) A confounder is an extraneous variable that is related to the exposure under investigation but that is also an independent risk factor for a disease. Confounding distorts the relationship between the exposure and the disease. For example, fatty food intake is a potential confounder in a study evaluating the association between physical activity and obesity, as people who do not exercise regularly may be more likely to consume a high-fat diet.

(Choice B) Lead-time bias refers to the apparent prolongation of survival in patients who undergo a screening test that allows for earlier diagnosis but does not actually improve prognosis. For example, a patient with pancreatic cancer presents with metastases at age 58 for which he receives treatment, but he dies at age 60. Had the same patient been screened and diagnosed with pancreatic cancer at age 55, he would have received earlier treatment, but the disease could have progressed such that he would have still died at age 60 (ie, age at death would not have been affected by screening).

(Choice D) Recall bias results from inaccurate recall of past exposure by people in a study; it applies most often to retrospective designs such as case-control studies. People who have experienced an adverse outcome are more likely to recall previous risk factors than people who have not experienced an adverse outcome.

(Choice E) Selection bias can occur with inappropriate (ie, nonrandom) selection methods or through selective attrition of the study participants. For example, many patients drop from a clinical trial due to the severity of side effects associated with the treatment. This type of attrition is selective (ie, different attrition rates between groups) and reduces the generalizability of the study.

Educational objective:
Observer bias occurs when the investigator's evaluation is affected by preconceived expectations or prior knowledge, typically leading to overestimation of the disease association or treatment effects. This type of bias can be reduced by conducting a blinded study in which observers

are more likely to be blinded to the study's objectives and the patients' medical history.

(Choice A) A confounder is an extraneous variable that is related to the exposure under investigation but that is also an independent risk factor for a disease. Confounding distorts the relationship between the exposure and the disease. For example, fatty food intake is a potential confounder in a study evaluating the association between physical activity and obesity, as people who do not exercise regularly may be more likely to consume a high-fat diet.

(Choice B) Lead-time bias refers to the apparent prolongation of survival in patients who undergo a screening test that allows for earlier diagnosis but does not actually improve prognosis. For example, a patient with pancreatic cancer presents with metastases at age 58 for which he receives treatment, but he dies at age 60. Had the same patient been screened and diagnosed with pancreatic cancer at age 55, he would have received earlier treatment, but the disease could have progressed such that he would have still died at age 60 (ie, age at death would not have been affected by screening).

(Choice D) Recall bias results from inaccurate recall of past exposure by people in a study; it applies most often to retrospective designs such as case-control studies. People who have experienced an adverse outcome are more likely to recall previous risk factors than people who have not experienced an adverse outcome.

(Choice E) Selection bias can occur with inappropriate (ie, nonrandom) selection methods or through selective attrition of the study participants. For example, many patients drop from a clinical trial due to the severity of side effects associated with the treatment. This type of attrition is selective (ie, different attrition rates between groups) and reduces the generalizability of the study.

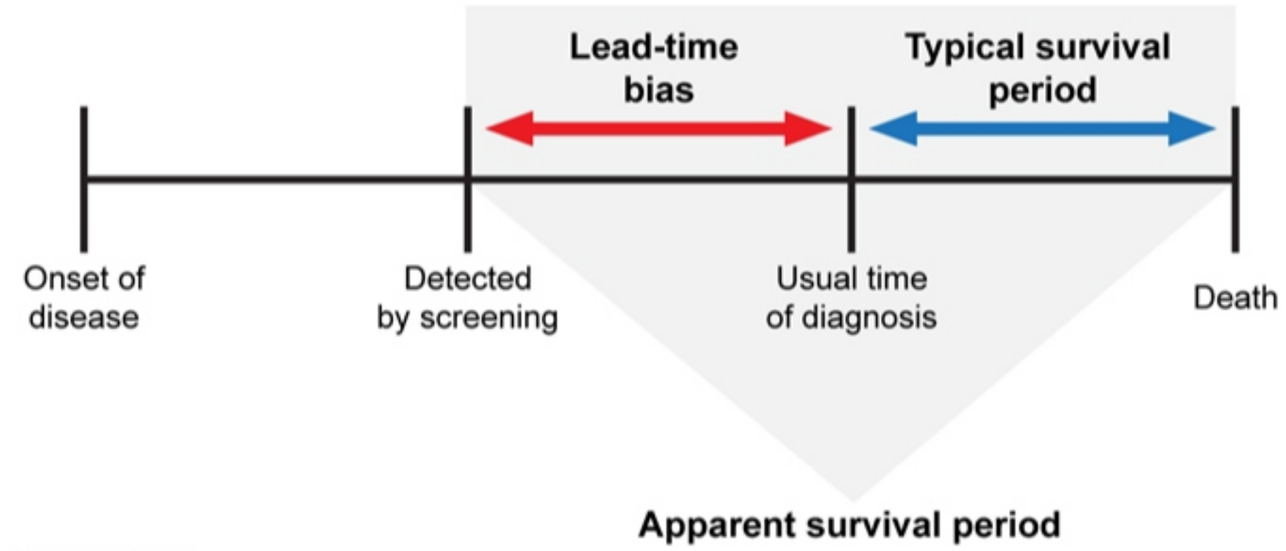
Educational objective:
Observer bias occurs when the investigator's evaluation is affected by preconceived expectations or prior knowledge, typically leading to overestimation of the disease association or treatment effects. This type of bias can be reduced by conducting a blinded study in which observers are unaware of study details and patient characteristics that could unduly influence them.

Biostatistics	Biostatistics & Epidemiology	Bias
Subject	System	Topic

Copyright © UWorld. All rights reserved.

are more likely to be blinded to the study's objectives and the patients' medical history.

Exhibit Display



Zoom In | Zoom Out | Reset | Add To New Card | Existing Card

Researchers want to explore the association between selective serotonin reuptake inhibitor (SSRI) use and pulmonary tuberculosis. A random sample of subjects age 20-84 with newly diagnosed pulmonary tuberculosis and an independent random sample of subjects without pulmonary tuberculosis are enrolled in the study. Subsequently, subjects who never had a prescription for an SSRI are defined as "never users," and those who ever had a prescription for an SSRI are defined as "users." The study compares the frequency of SSRI use in subjects with and without a diagnosis of pulmonary tuberculosis and determines that SSRI use is not associated with pulmonary tuberculosis. Which of the following best describes the design of this study?

A. Case-control study

B. Cross-sectional study

C. Prospective cohort study

D. Randomized controlled trial

E. Retrospective cohort study

Submit

Researchers want to explore the association between selective serotonin reuptake inhibitor (SSRI) use and pulmonary tuberculosis. A random sample of subjects age 20-84 with newly diagnosed pulmonary tuberculosis and an independent random sample of subjects without pulmonary tuberculosis are enrolled in the study. Subsequently, subjects who never had a prescription for an SSRI are defined as "never users," and those who ever had a prescription for an SSRI are defined as "users." The study compares the frequency of SSRI use in subjects with and without a diagnosis of pulmonary tuberculosis and determines that SSRI use is not associated with pulmonary tuberculosis. Which of the following best describes the design of this study?

- ✓ A. Case-control study (62%)
- B. Cross-sectional study (13%)
- C. Prospective cohort study (2%)
- D. Randomized controlled trial (2%)
- E. Retrospective cohort study (18%)

Omitted
Correct answer
A

62%
Answered correctly

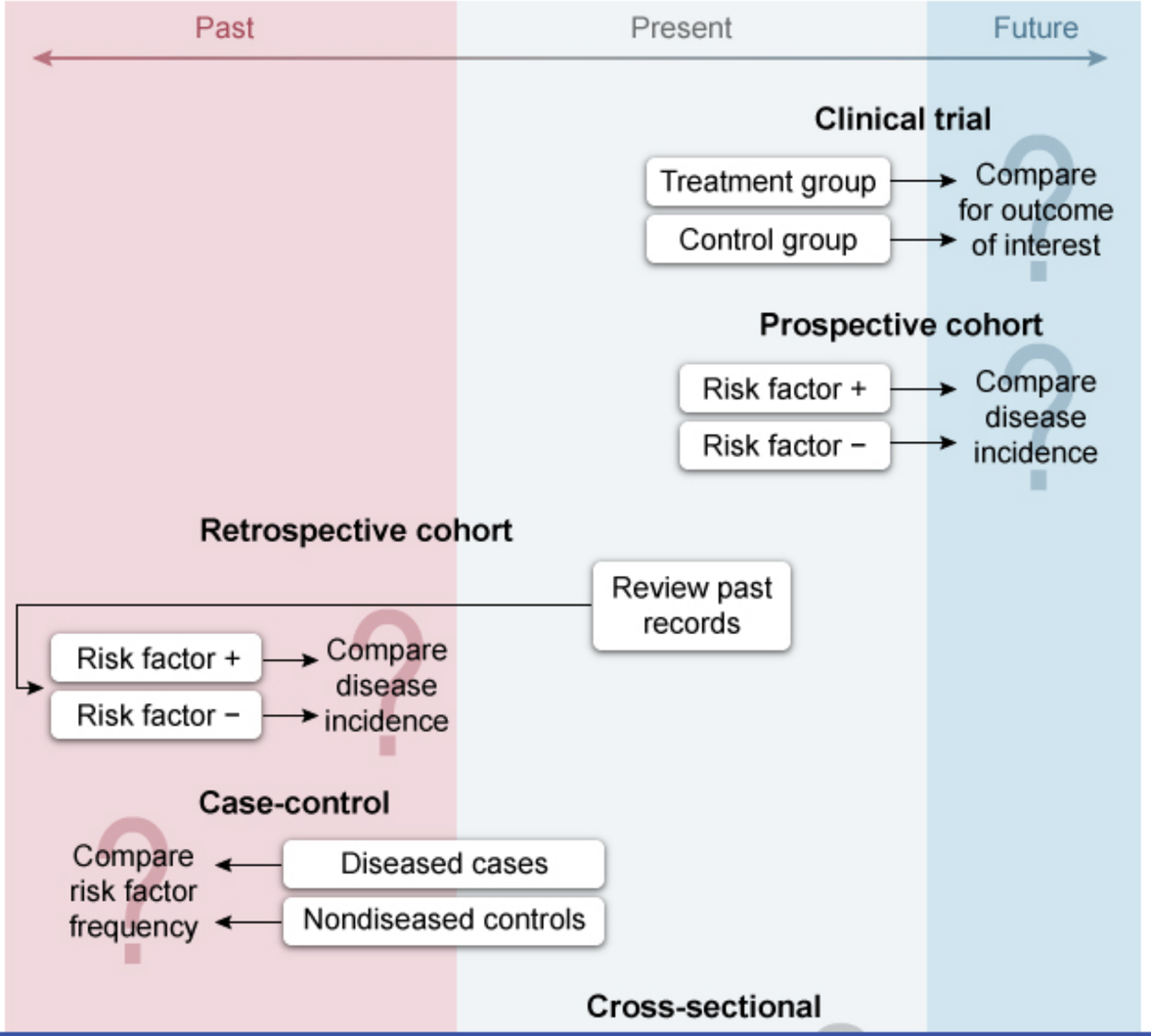
02 secs
Time Spent

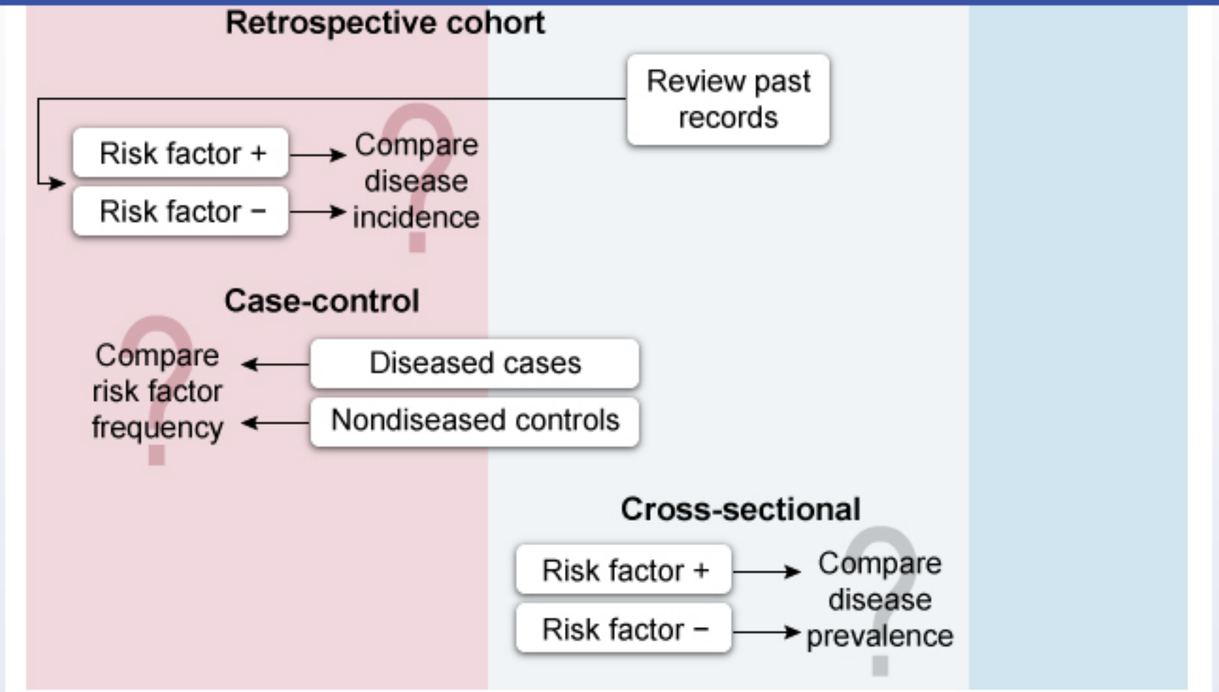
05/13/2020
Last Updated

Explanation

Temporality of different study designs

Temporality of different study designs





©UWorld

Research studies are broadly classified as having an observational or experimental design depending on the control the researcher has over the independent variables (eg, exposure to risk factors, treatments).

- In **observational studies** (eg, case series, case-control studies, cohort studies, cross-sectional studies), the researcher has **no control** over the independent variables (eg, exposures, risk factors).
- In **experimental studies** (eg, randomized controlled trials), the researcher **controls** and randomly assigns the independent variables (eg, treatments, interventions).

A **case-control** study is an observational design in which potential participants are initially identified as **cases** or **controls** according to the **dependent variable or outcome** (eg, disease of interest). In the study described above, subjects age 20-84 newly diagnosed with pulmonary tuberculosis are cases, and subjects without pulmonary tuberculosis are controls. Once cases and controls are identified, the presence of past

Research studies are broadly classified as having an observational or experimental design depending on the control the researcher has over the independent variables (eg, exposure to risk factors, treatments).

- In **observational studies** (eg, case series, case-control studies, cohort studies, cross-sectional studies), the researcher has **no control** over the independent variables (eg, exposures, risk factors).
- In **experimental studies** (eg, randomized controlled trials), the researcher **controls** and randomly assigns the independent variables (eg, treatments, interventions).

A **case-control** study is an observational design in which potential participants are initially identified as **cases** or **controls** according to the **dependent variable or outcome** (eg, disease of interest). In the study described above, subjects age 20-84 newly diagnosed with pulmonary tuberculosis are cases, and subjects without pulmonary tuberculosis are controls. Once cases and controls are identified, the presence of past exposure to ≥1 **risk factors** of interest is determined in each group. In this study, the risk factor is selective serotonin reuptake inhibitor (SSRI) use ("never users" are subjects who never had a prescription for an SSRI, and "users" are subjects who ever had a prescription for an SSRI). Finally, the **frequency** of exposure to the risk factor is compared between cases and controls to estimate the **association** between the risk factor and the outcomes. Conducted as designed, this study can determine if SSRI use is associated with pulmonary tuberculosis.

(Choice B) In contrast to participants in a case-control study, who are selected based on outcome status (eg, presence or absence of tuberculosis), participants in a cross-sectional study are randomly selected from a population of interest (eg, random sample from a region with high tuberculosis prevalence) and then simultaneously categorized according to risk factor (eg, SSRI use) and outcome (eg, tuberculosis) status. The goal is to determine prevalence by providing a snapshot of the population.

(Choices C and E) In a cohort study, 2 groups of individuals (ie, cohorts) are initially identified based on their exposure status to a specific risk factor (eg, SSRI use) rather than based on their outcome status (eg, pulmonary tuberculosis), as seen in this example. These 2 cohorts are then followed over time (ie, prospectively) to assess development of the outcome. Sometimes the exposure status is determined retrospectively, typically using medical records, and patients are tracked from the point of exposure onward.

(Choice D) A randomized controlled trial follows individuals who have been randomized to either a treatment arm or a control arm and compares the effect of the intervention.

Educational objective:
A case-control study is an observational study design; it begins with individuals who have the outcome (cases) and compares them with

dependent variable or outcome (eg, disease of interest). In the study described above, subjects age 20-81 newly diagnosed with pulmonary tuberculosis are cases, and subjects without pulmonary tuberculosis are controls. Once cases and controls are identified, the presence of past exposure to ≥1 **risk factors** of interest is determined in each group. In this study, the risk factor is selective serotonin reuptake inhibitor (SSRI) use ("never users" are subjects who never had a prescription for an SSRI, and "users" are subjects who ever had a prescription for an SSRI). Finally, the **frequency** of exposure to the risk factor is compared between cases and controls to estimate the **association** between the risk factor and the outcomes. Conducted as designed, this study can determine if SSRI use is associated with pulmonary tuberculosis.

(Choice B) In contrast to participants in a case-control study, who are selected based on outcome status (eg, presence or absence of tuberculosis), participants in a cross-sectional study are randomly selected from a population of interest (eg, random sample from a region with high tuberculosis prevalence) and then simultaneously categorized according to risk factor (eg, SSRI use) and outcome (eg, tuberculosis) status. The goal is to determine prevalence by providing a snapshot of the population.

(Choices C and E) In a cohort study, 2 groups of individuals (ie, cohorts) are initially identified based on their exposure status to a specific risk factor (eg, SSRI use) rather than based on their outcome status (eg, pulmonary tuberculosis), as seen in this example. These 2 cohorts are then followed over time (ie, prospectively) to assess development of the outcome. Sometimes the exposure status is determined retrospectively, typically using medical records, and patients are tracked from the point of exposure onward.

(Choice D) A randomized controlled trial follows individuals who have been randomized to either a treatment arm or a control arm and compares the effect of the intervention.

Educational objective:
A case-control study is an observational study design; it begins with individuals who have the outcome (cases) and compares them with individuals who do not have the outcome (controls) according to history of exposure to ≥1 risk factors.

References

- Epidemiology in practice: case-control studies.

Biostatistics	Biostatistics & Epidemiology	Study designs
Subject	System	Topic

Copyright © UWorld. All rights reserved.

Researchers at an academic trauma center conducted a randomized clinical trial comparing 2 surgical techniques (Technique A and Technique B) for repairing a mandible fracture. They examined rates of infectious complications and rates of malocclusion (defined as misaligned dental arches on visual examination when the jaw is closed). A total of 100 patients were enrolled in the study. Outcome assessments of each patient were conducted at 2 weeks, 4 weeks, and 8 weeks after surgery. Results showed that the relative rate of infection with Technique A compared with Technique B was 0.86 with a 95% confidence interval of 0.57-1.28; the 2 techniques had similar malocclusion rates at 8 weeks after surgery. Which of the following statements most accurately represents the comparison of Technique A and Technique B in a clinical care setting?

A. Neither surgery technique is superior

B. Technique A is superior to Technique B

C. Technique B is superior to Technique A

D. The techniques should not be used in a clinical care setting

Submit

Researchers at an academic trauma center conducted a randomized clinical trial comparing 2 surgical techniques (Technique A and Technique B) for repairing a mandible fracture. They examined rates of infectious complications and rates of malocclusion (defined as misaligned dental arches on visual examination when the jaw is closed). A total of 100 patients were enrolled in the study. Outcome assessments of each patient were conducted at 2 weeks, 4 weeks, and 8 weeks after surgery. Results showed that the relative rate of infection with Technique A compared with Technique B was 0.86 with a 95% confidence interval of 0.57-1.28; the 2 techniques had similar malocclusion rates at 8 weeks after surgery. Which of the following statements most accurately represents the comparison of Technique A and Technique B in a clinical care setting?

- ✓ A. Neither surgery technique is superior (75%)
- B. Technique A is superior to Technique B (12%)
- C. Technique B is superior to Technique A (8%)
- D. The techniques should not be used in a clinical care setting (3%)

Omitted
Correct answer
A

75%
Answered correctly

02 secs
Time Spent

06/03/2020
Last Updated

Explanation

The **relative risk (RR)** is commonly used in cohort and experimental studies as a measure of **association** between a risk factor or **treatment** and a disease or **outcome**. RR is equal to the risk of an outcome or disease (eg, infection) in a group exposed to a treatment (eg, Treatment A) divided by the risk of the same outcome or disease in a group exposed to a different treatment (eg, Treatment B).

RR = (risk of disease in Treatment A) / (risk of disease in Treatment B)

The RR may be interpreted as follows:

The **relative risk (RR)** is commonly used in cohort and experimental studies as a measure of **association** between a risk factor or **treatment** and a disease or **outcome**. RR is equal to the risk of an outcome or disease (eg, infection) in a group exposed to a treatment (eg, Treatment A) divided by the risk of the same outcome or disease in a group exposed to a different treatment (eg, Treatment B).

RR = (risk of disease in Treatment A) / (risk of disease in Treatment B)

The RR may be interpreted as follows:

- RR < 1.0 indicates that exposure to Treatment A is associated with a decrease in risk of disease with respect to Treatment B.
- **RR = 1.0 (null value)** indicates that there is no association between the treatments and risk of disease.
- RR > 1.0 indicates that exposure to Treatment A is associated with an increase in risk of disease with respect to Treatment B.

The **confidence interval** (CI) of a RR indicates, with a certain confidence level, whether a given RR is statistically different from the null value (ie, RR = 1). A CI that **includes the null value** (ie, RR = 1) is **not statistically significant**, and a CI that **excludes** the null value (ie, RR = 1) is **statistically significant** (at that confidence level).

In this study, the RR for infection among those assigned to surgery Technique A compared to those assigned to surgery Technique B is 0.86 with a 95% CI of 0.57-1.28. An RR = 0.86 indicates that exposure to surgery Technique A decreases the risk of infection. However, because the CI includes the null value (ie, RR = 1), the decrease in risk is **not statistically significant**. Therefore, neither surgery technique is superior at reducing infection **(Choices B and C)**.

(Choice D) The study results show that the surgery techniques have similar rates of infectious complications and malocclusion. Assuming these rates are acceptable and there are no other major complications, there is no reason to determine the surgical techniques should not be used in a clinical care setting, but further information may be useful.

Educational objective:

- The relative risk (RR) is a measure of association: RR < 1.0 indicates a lower risk of disease in the exposed group relative to the nonexposed group; RR = 1.0 indicates no association between exposure and disease; and RR > 1.0 indicates a greater risk of disease in the exposed group relative to the nonexposed group.

- RR < 1.0 indicates that exposure to Treatment A is associated with a decrease in risk of disease with respect to Treatment B.
- **RR = 1.0 (null value)** indicates that there is no association between the treatments and risk of disease.
- RR > 1.0 indicates that exposure to Treatment A is associated with an increase in risk of disease with respect to Treatment B.

The **confidence interval** (CI) of a RR indicates, with a certain confidence level, whether a given RR is statistically different from the null value (ie, RR = 1). A CI that **includes the null value** (ie, RR = 1) is **not statistically significant**, and a CI that **excludes** the null value (ie, RR = 1) is **statistically significant** (at that confidence level).

In this study, the RR for infection among those assigned to surgery Technique A compared to those assigned to surgery Technique B is 0.86 with a 95% CI of 0.57-1.28. An RR = 0.86 indicates that exposure to surgery Technique A decreases the risk of infection. However, because the CI includes the null value (ie, RR = 1), the decrease in risk is **not statistically significant**. Therefore, neither surgery technique is superior at reducing infection **(Choices B and C)**.

(Choice D) The study results show that the surgery techniques have similar rates of infectious complications and malocclusion. Assuming these rates are acceptable and there are no other major complications, there is no reason to determine the surgical techniques should not be used in a clinical care setting, but further information may be useful.

Educational objective:

- The relative risk (RR) is a measure of association: RR < 1.0 indicates a lower risk of disease in the exposed group relative to the nonexposed group; RR = 1.0 indicates no association between exposure and disease; and RR > 1.0 indicates a greater risk of disease in the exposed group relative to the nonexposed group.
- A confidence interval (CI) that includes the null value for an RR (ie, RR = 1) is not statistically significant, and a CI that excludes the null value (ie, RR = 1) is statistically significant.

Biostatistics	Biostatistics & Epidemiology	Statistical vs. clinical significance
Subject	System	Topic

Copyright © UWorld. All rights reserved.

A new serologic test has been developed for the detection of active pulmonary tuberculosis and is compared to the gold standard of sputum mycobacterial culture. A total of 1000 subjects are randomly selected for testing from a population with a high prevalence of tuberculosis. Results of the study are given below:

	Sputum culture positive	Sputum culture negative	
Serologic test positive	130	60	190
Serologic test negative	50	760	810
	180	820	1000

Which of the following is the positive predictive value of the test under study?

A. 130/180
B. 130/190
C. 180/1000
D. 760/810
E. 760/820

Submit

A new serologic test has been developed for the detection of active pulmonary tuberculosis and is compared to the gold standard of sputum mycobacterial culture. A total of 1000 subjects are randomly selected for testing from a population with a high prevalence of tuberculosis. Results of the study are given below:

	Sputum culture positive	Sputum culture negative	
Serologic test positive	130	60	190
Serologic test negative	50	760	810
	180	820	1000

Which of the following is the positive predictive value of the test under study?

A. 130/180 (11%)
B. 130/190 (86%)
C. 180/1000 (1%)
D. 760/810 (0%)
E. 760/820 (0%)

Omitted
Correct answer
B

86%
Answered correctly

02 secs
Time Spent

01/28/2020
Last Updated

Explanation

The **positive predictive value (PPV)** of a test answers the question: If the test result is positive, what is the probability that a patient has the disease? PPV is calculated as the proportion of subjects who truly have the disease among all those with a positive test result. In contrast, the **negative predictive value (NPV)** of a test answers the question: If the test result is negative, what is the probability that a patient does not have the disease? It is calculated as the proportion of subjects who are truly free of disease among all those with a negative test result. Predictive values are of prime importance to physicians because, in clinical practice, patients will present more often with a positive or negative test than with a defined diseased or disease-free state.

Consider the following 2 × 2 (contingency) table:

	Positive condition	**Negative condition**	
Positive test result	TP	FP	PPV = TP/(TP + FP)
Negative test result	FN	TN	NPV = TN/(TN + FN)
	Sensitivity = TP/(TP + FN)	Specificity = TN/(TN + FP)	

FN = false negative; **FP** = false positive; **NPV** = negative predictive value; **PPV** = positive predictive value; **TN** = true negative; **TP** = true positive.

PPV = TP / (TP + FP), where TP is true positives and FP is false positives.

NPV = TN / (TN + FN), where TN is true negatives and FN is false negatives.

In this case, PPV = 130 / (130 + 60) = 130/190.

(Choices A and E) Sensitivity is a test's ability to correctly identify individuals with the disease. Sensitivity = TP / (TP + FN) = 130 / (130 + 50) = 130/180. Specificity is a test's ability to correctly identify individuals without the disease. Specificity = TN / (TN + FP) = 760 / (760 + 60) = 760/820.

	Positive condition	Negative condition	
Positive test result	TP	FP	PPV = TP/(TP + FP)
Negative test result	FN	TN	NPV = TN/(TN + FN)
	Sensitivity = TP/(TP + FN)	Specificity = TN/(TN + FP)	

FN = false negative; **FP** = false positive; **NPV** = negative predictive value; **PPV** = positive predictive value; **TN** = true negative; **TP** = true positive.

PPV = TP / (TP + FP), where TP is true positives and FP is false positives.

NPV = TN / (TN + FN), where TN is true negatives and FN is false negatives.

In this case, PPV = 130 / (130 + 60) = 130/190.

(Choices A and E) Sensitivity is a test's ability to correctly identify individuals with the disease. Sensitivity = TP / (TP + FN) = 130 / (130 + 50) = 130/180. Specificity is a test's ability to correctly identify individuals without the disease. Specificity = TN / (TN + FP) = 760 / (760 + 60) = 760/820.

(Choice C) The total number of individuals who have tuberculosis (based on the gold standard of positive sputum culture) is 180. The total number of individuals in this sample is 1000. Therefore, the prevalence of tuberculosis in this sample is 180/1000.

(Choice D) NPV = 760 / (760 + 50) = 760/810.

Educational objective:
The positive predictive value (PPV) of a test answers the question: If the test result is positive, what is the probability that a patient has the disease? PPV = true positives / (true positives + false positives).

Negative test result	FN	TN	NPV = TN/(TN + FN)
	Sensitivity = TP/(TP + FN)	Specificity = TN/(TN + FP)	

FN = false negative; **FP** = false positive; **NPV** = negative predictive value; **PPV** = positive predictive value; **TN** = true negative; **TP** = true positive.

PPV = TP / (TP + FP), where TP is true positives and FP is false positives.

NPV = TN / (TN + FN), where TN is true negatives and FN is false negatives.

In this case, PPV = 130 / (130 + 60) = 130/190.

(Choices A and E) Sensitivity is a test's ability to correctly identify individuals with the disease. Sensitivity = TP / (TP + FN) = 130 / (130 + 50) = 130/180. Specificity is a test's ability to correctly identify individuals without the disease. Specificity = TN / (TN + FP) = 760 / (760 + 60) = 760/820.

(Choice C) The total number of individuals who have tuberculosis (based on the gold standard of positive sputum culture) is 180. The total number of individuals in this sample is 1000. Therefore, the prevalence of tuberculosis in this sample is 180/1000.

(Choice D) NPV = 760 / (760 + 50) = 760/810.

Educational objective:
The positive predictive value (PPV) of a test answers the question: If the test result is positive, what is the probability that a patient has the disease? PPV = true positives / (true positives + false positives).

Biostatistics	Biostatistics & Epidemiology	Sensitivity, specificity, positive predictive value PPV, negative predictive value NPV
Subject	System	Topic

Copyright © UWorld. All rights reserved.

A prospective study evaluates the relationship between regular antioxidant supplement use (vitamins C and E) and the risk of stroke in healthy and physically active men age 40-60. The study compares the risk of stroke among men who consumed antioxidant supplements for ≥5 years and among men who consumed antioxidant supplements for <5 years, as compared to a reference group of men who never consumed antioxidant supplements. According to the study results, men who consumed antioxidant supplements for <5 years and men who consumed antioxidant supplements for ≥5 years have stroke relative risks of 0.95 ($p = 0.45$) and 0.75 ($p < 0.01$), respectively, when compared to the reference group. The results of the study were adjusted to account for baseline differences related to healthy behaviors and overall health. Which of the following factors most likely explains why the relative risk of stroke is lower with longer antioxidant use?

A. Accumulation effect

B. Lead-time bias

C. Observer bias

D. Rare disease assumption

E. Selection bias

Submit

A prospective study evaluates the relationship between regular antioxidant supplement use (vitamins C and E) and the risk of stroke in healthy and physically active men age 40-60. The study compares the risk of stroke among men who consumed antioxidant supplements for ≥5 years and among men who consumed antioxidant supplements for <5 years, as compared to a reference group of men who never consumed antioxidant supplements. According to the study results, men who consumed antioxidant supplements for <5 years and men who consumed antioxidant supplements for ≥5 years have stroke relative risks of 0.95 ($p = 0.45$) and 0.75 ($p < 0.01$), respectively, when compared to the reference group. The results of the study were adjusted to account for baseline differences related to healthy behaviors and overall health. Which of the following factors most likely explains why the relative risk of stroke is lower with longer antioxidant use?

- ✓ A. Accumulation effect (44%)
- B. Lead-time bias (27%)
- C. Observer bias (5%)
- D. Rare disease assumption (2%)
- E. Selection bias (20%)

Omitted
Correct answer
A

44%
Answered correctly

02 secs
Time Spent

05/18/2020
Last Updated

Explanation

Antioxidant use for **≥5 years** was associated with a **statistically significant** ($p < 0.01$) decreased risk of stroke (relative risk [RR] = 0.75 <1). By contrast, antioxidant use for **<5 years** demonstrated a **nonstatistically significant** ($p > 0.05$) smaller reduction in risk (RR = 0.95); in other words, antioxidant use for <5 years seems to have no benefit on stroke risk reduction. This discrepancy can be explained if reduction in stroke risk is

Antioxidant use for **≥5 years** was associated with a **statistically significant** ($p < 0.01$) decreased risk of stroke (relative risk [RR] = 0.75 <1). By contrast, antioxidant use for **<5 years** demonstrated a **nonstatistically significant** ($p > 0.05$) smaller reduction in risk (RR = 0.95); in other words, antioxidant use for <5 years seems to have no benefit on stroke risk reduction. This discrepancy can be explained if reduction in stroke risk is associated with the **duration of exposure** to antioxidant supplementation; in other words, a minimum cumulative exposure to antioxidants may be necessary for the exposure to have a significant effect on stroke risk.

This concept of **accumulation effect** can apply to both risk factors and risk reducers. The effect of exposure to risk factors may depend on the duration and intensity of the exposure; long-term exposure may be necessary well before an effect on the disease process is clinically evident (eg, lung cancer developing after decades of smoking exposure). Similarly, exposure to certain **risk reducers** must occur continuously over extended periods before disease outcome is affected. In this case, ≥5 years of continuous antioxidant use (risk reducer) were required to reveal their protective effect on stroke.

(Choice B) Lead time is the time between the initial detection of a disease and a specific outcome or measured endpoint. Lead-time bias can occur when a test detects or diagnoses the disease at an earlier stage than another test without impacting the natural history of the disease. A study comparing disease survival times may then erroneously conclude that using the earlier-detection test prolongs survival, when in actuality the increased survival time is due solely to earlier detection of the disease.

(Choice C) Observer bias occurs when an observer misclassifies data due to individual differences in interpretation or preconceived expectations regarding a study. It can be reduced by performing a double-blind study (ie, both observers and participants unaware of randomized parameters) and by having multiple observers encode and verify the recorded data.

(Choice D) Diseases with a very low incidence rate in the population will also have a very low prevalence. According to the rare disease assumption, the odds ratio approximates RR when disease prevalence is low (eg, <10%).

(Choice E) Selection bias can occur with inappropriate (ie, nonrandom) selection methods or through selective attrition of the study participants. Although selection bias is a possible limitation of the study (eg, individuals who take vitamins daily may be more likely to be already healthy), the fact that the study was limited to healthy, physically active men and that it adjusted for baseline differences should considerably reduce the effect of this potential bias.

Educational objective:
The concept of accumulation effect can be applied to disease pathogenesis and exposure to risk modifiers. Cumulative exposure to a risk factor

duration and intensity of the exposure; long-term exposure may be necessary well before an effect on the disease process is clinically evident (eg, lung cancer developing after decades of smoking exposure). Similarly, exposure to certain **risk reducers** must occur continuously over extended periods before disease outcome is affected. In this case, ≥5 years of continuous antioxidant use (risk reducer) were required to reveal their protective effect on stroke.

(Choice B) Lead time is the time between the initial detection of a disease and a specific outcome or measured endpoint. Lead-time bias can occur when a test detects or diagnoses the disease at an earlier stage than another test without impacting the natural history of the disease. A study comparing disease survival times may then erroneously conclude that using the earlier-detection test prolongs survival, when in actuality the increased survival time is due solely to earlier detection of the disease.

(Choice C) Observer bias occurs when an observer misclassifies data due to individual differences in interpretation or preconceived expectations regarding a study. It can be reduced by performing a double-blind study (ie, both observers and participants unaware of randomized parameters) and by having multiple observers encode and verify the recorded data.

(Choice D) Diseases with a very low incidence rate in the population will also have a very low prevalence. According to the rare disease assumption, the odds ratio approximates RR when disease prevalence is low (eg, <10%).

(Choice E) Selection bias can occur with inappropriate (ie, nonrandom) selection methods or through selective attrition of the study participants. Although selection bias is a possible limitation of the study (eg, individuals who take vitamins daily may be more likely to be already healthy), the fact that the study was limited to healthy, physically active men and that it adjusted for baseline differences should considerably reduce the effect of this potential bias.

Educational objective:
The concept of accumulation effect can be applied to disease pathogenesis and exposure to risk modifiers. Cumulative exposure to a risk factor or risk reducer must sometimes occur for prolonged periods before a clinically significant effect is detected.

Biostatistics	Biostatistics & Epidemiology	Risk
Subject	System	Topic

Copyright © UWorld. All rights reserved.

duration and intensity of the exposure; long-term exposure may be necessary well before an effect on the disease process is clinically evident (eg,

Exhibit Display

Relative risk ≈ Odds ratio

$$RR = \frac{a}{a+b} / \frac{c}{c+d} \approx \frac{a}{b} / \frac{c}{d} = \frac{ad}{bc} = OR$$

If a <<b & c <<d (low disease incidence)

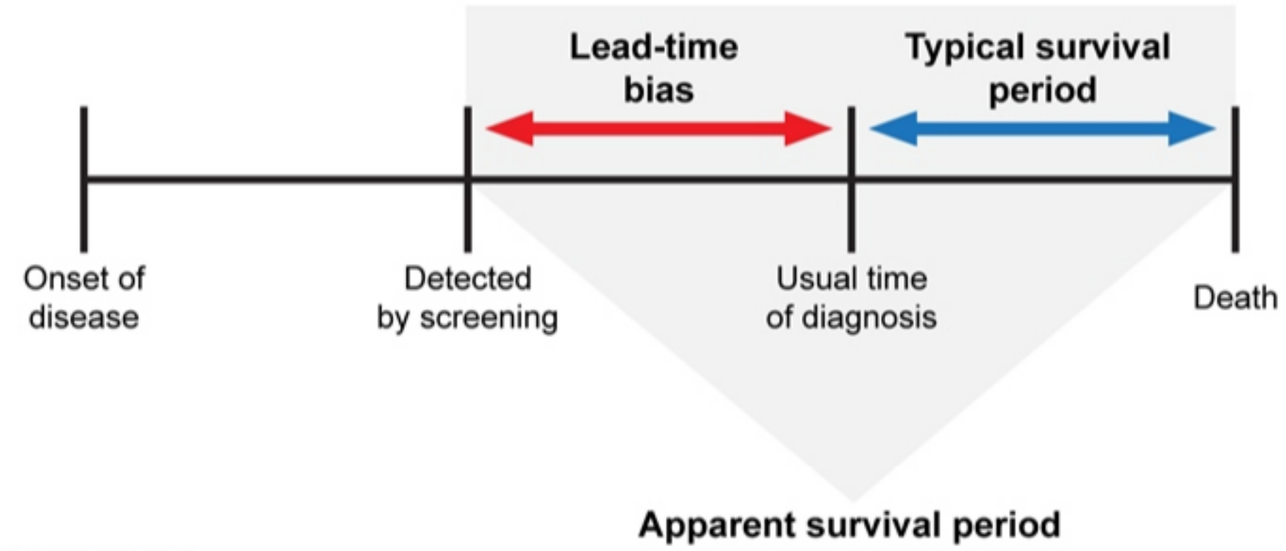
OR = odds ratio; **RR** = relative risk

©UWorld

Zoom In | Zoom Out | Reset | Add To New Card | Existing Card

duration and intensity of the exposure; long-term exposure may be necessary well before an effect on the disease process is clinically evident (eg,

Exhibit Display



Zoom In | Zoom Out | Reset | Add To New Card | Existing Card

A study examined the effect of a low-carbohydrate diet compared to a low-fat diet on body weight. A sample of 150 overweight but otherwise healthy adults from a large city was enrolled in the study and randomly assigned in a 1:1 ratio to either the low-carbohydrate (40 g/d) or the low-fat (<7% saturated fat) diet. At 12 months, a greater body weight change was reported in the low-carbohydrate diet group compared to the low-fat diet group, with a mean difference in body weight change of −3.5 kg ($p = 0.01$, predetermined significance level = 0.05). Which of the following is the most accurate interpretation of the results of this study?

A. The observed mean difference in body weight change of −3.5 kg is not statistically significant

B. The probability of observing a mean difference in body weight change of −3.5 kg is 0.01

C. There is a 1% chance of observing a mean difference in body weight change of at least −3.5 kg when no difference between groups is assumed

D. There is a 1% chance that an adult on a low-carbohydrate diet will have a body weight change of at least −3.5 kg at 12 months after starting the diet

E. There is a 1% chance that the mean difference in body weight change is biased in favor of the low-carbohydrate diet group

Submit

A study examined the effect of a low-carbohydrate diet compared to a low-fat diet on body weight. A sample of 150 overweight but otherwise healthy adults from a large city was enrolled in the study and randomly assigned in a 1:1 ratio to either the low-carbohydrate (40 g/d) or the low-fat (<7% saturated fat) diet. At 12 months, a greater body weight change was reported in the low-carbohydrate diet group compared to the low-fat diet group, with a mean difference in body weight change of −3.5 kg (p = 0.01, predetermined significance level = 0.05). Which of the following is the most accurate interpretation of the results of this study?

A. The observed mean difference in body weight change of −3.5 kg is not statistically significant (3%)

B. The probability of observing a mean difference in body weight change of −3.5 kg is 0.01 (5%)

✓ C. There is a 1% chance of observing a mean difference in body weight change of at least −3.5 kg when no difference between groups is assumed (57%)

D. There is a 1% chance that an adult on a low-carbohydrate diet will have a body weight change of at least −3.5 kg at 12 months after starting the diet (9%)

E. There is a 1% chance that the mean difference in body weight change is biased in favor of the low-carbohydrate diet group (24%)

Omitted
Correct answer
C

57%
Answered correctly

02 secs
Time Spent

06/22/2020
Last Updated

Explanation

Statistical tests contrast a null hypothesis (H_0) and an alternative hypothesis (H_a), in this case:

- H_0 is a claim of **no difference** between populations (eg, no difference in mean body weight between the low-carbohydrate population and the low-fat population).

Explanation

Statistical tests contrast a null hypothesis (H_0) and an alternative hypothesis (H_a), in this case:

- H_0 is a claim of **no difference** between populations (eg, no difference in mean body weight between the low-carbohydrate population and the low-fat population).
- H_a is a claim of a **difference** between populations (eg, difference in mean body weight between the 2 populations).

Statistical inference uses data from **samples** (eg, 75 adults in the low-carbohydrate sample, 75 adults in the low-fat sample) to draw conclusions about underlying **populations**. Sample estimates or differences (eg, sample mean, difference in mean between 2 samples) generally vary with distinct samples (eg, if a different sample of 75 adults was chosen for each group) and may be close but not equal to the underlying population value.

One way to account for **sampling variation** is to calculate the ***p*-value**, which is the **probability of obtaining a sample value at least as large as the one observed** when the population value claimed in **H_0 is assumed to be true**. The magnitude of the *p*-value compared to a predetermined significance level (eg, 0.05 [or 5%] commonly used as the threshold) determines whether there is convincing evidence against H_0.

- A low *p*-value (eg, <0.05) occurs when the sample value significantly disagrees with the population value claimed in H_0 and provides convincing evidence against H_0 (ie, H_0 is probably wrong). Results are considered statistically significant.
- A high *p*-value (eg, ≥0.05) occurs when the sample value is close to the population value claimed in H_0 and provides convincing evidence in favor of H_0 (ie, H_0 might be correct). Results are considered not statistically significant.

In this case, the given ***p*-value = 0.01**; therefore, there is a **1% chance** (ie, 0.01) of observing a mean difference in body weight change (ie, sample estimate) of **at least −3.5 kg** between the low-fat and the low-carbohydrate samples when **no difference between the populations is assumed** (ie, H_0 is assumed to be true) **(Choice B)**.

(Choices A, D, and E) The *p*-value of 0.01 is less than 0.05; therefore, the observed mean difference is statistically significant at the 5% level. The *p*-value is not associated with individual observations in a sample. It accounts for sampling variation (ie, random variation) not bias (ie, systematic variation).

Educational objective:

The *p*-value is the probability of obtaining a result (ie, sample estimate) at least as large as the one observed when the population value claimed in

Statistical inference uses data from **samples** (eg, 75 adults in the low-carbohydrate sample, 75 adults in the low-fat sample) to draw conclusions about underlying **populations**. Sample estimates or differences (eg, sample mean, difference in mean between 2 samples) generally vary with distinct samples (eg, if a different sample of 75 adults was chosen for each group) and may be close but not equal to the underlying population value.

One way to account for **sampling variation** is to calculate the ***p*-value**, which is the **probability of obtaining a sample value at least as large as the one observed** when the population value claimed in **H_0 is assumed to be true**. The magnitude of the *p*-value compared to a predetermined significance level (eg, 0.05 [or 5%] commonly used as the threshold) determines whether there is convincing evidence against H_0.

- A low *p*-value (eg, <0.05) occurs when the sample value significantly disagrees with the population value claimed in H_0 and provides convincing evidence against H_0 (ie, H_0 is probably wrong). Results are considered statistically significant.
- A high *p*-value (eg, ≥0.05) occurs when the sample value is close to the population value claimed in H_0 and provides convincing evidence in favor of H_0 (ie, H_0 might be correct). Results are considered not statistically significant.

In this case, the given ***p*-value = 0.01**; therefore, there is a **1% chance** (ie, 0.01) of observing a mean difference in body weight change (ie, sample estimate) of **at least −3.5 kg** between the low-fat and the low-carbohydrate samples when **no difference between the populations is assumed** (ie, H_0 is assumed to be true) **(Choice B)**.

(Choices A, D, and E) The *p*-value of 0.01 is less than 0.05; therefore, the observed mean difference is statistically significant at the 5% level. The *p*-value is not associated with individual observations in a sample. It accounts for sampling variation (ie, random variation) not bias (ie, systematic variation).

Educational objective:
The *p*-value is the probability of obtaining a result (ie, sample estimate) at least as large as the one observed when the population value claimed in the null hypothesis is assumed to be true. A *p*-value <0.05 typically indicates that results are statistically significant.

Biostatistics	Biostatistics & Epidemiology	P-value and confidence interval
Subject	System	Topic

Copyright © UWorld. All rights reserved.

Blood pressure measurements are obtained from a sample of individuals with no known medical conditions. For systolic blood pressure (SBP), the mean measurements and associated standard deviations (SDs) are shown by age group for men and women:

	Men	Women
Age group	Mean SBP (mm Hg) ± SD	Mean SBP (mm Hg) ± SD
35-44	120 ± 20	124 ± 18
45-54	131 ± 21	137 ± 24
55-64	141 ± 19	140 ± 20

If hypertension is defined as SBP >140 mm Hg, approximately what percentage of men age 35-44 in this sample will be classified as having hypertension, assuming a normal (Gaussian) distribution?

- A. 16%
- B. 34%
- C. 50%
- D. 68%
- E. 84%
- F. 95%

Submit

Blood pressure measurements are obtained from a sample of individuals with no known medical conditions. For systolic blood pressure (SBP), the mean measurements and associated standard deviations (SDs) are shown by age group for men and women:

	Men	Women
Age group	Mean SBP (mm Hg) ± SD	Mean SBP (mm Hg) ± SD
35-44	120 ± 20	124 ± 18
45-54	131 ± 21	137 ± 24
55-64	141 ± 19	140 ± 20

If hypertension is defined as SBP >140 mm Hg, approximately what percentage of men age 35-44 in this sample will be classified as having hypertension, assuming a normal (Gaussian) distribution?

- ✓ A. 16% (83%)
- B. 34% (8%)
- C. 50% (1%)
- D. 68% (2%)
- E. 84% (1%)
- F. 95% (2%)

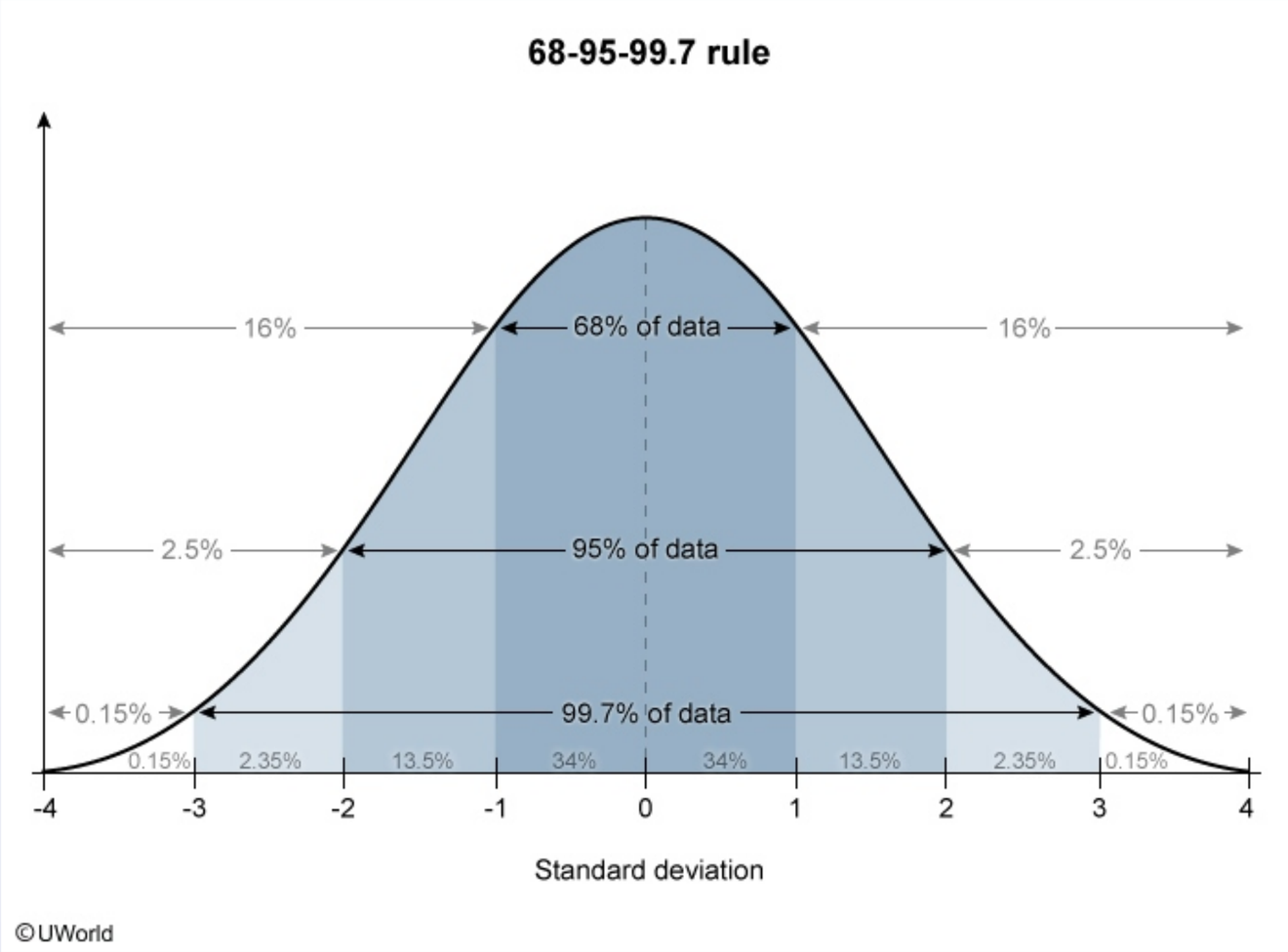
Omitted
Correct answer
A

83%
Answered correctly

02 secs
Time Spent

05/28/2020
Last Updated

Explanation



A **normal (Gaussian) distribution** refers to a symmetrical, bell-shaped distribution with a fixed proportion of observations lying within a certain distance of the mean. This distance is called the **standard deviation (SD)** and reflects the **degree of dispersion** from the mean. According to the properties of this distribution, 68% of observations lie within 1 SD on either side of the mean, with half (ie, 68/2 = 34%) above and half (34%)

©UWorld

A **normal (Gaussian) distribution** refers to a symmetrical, bell-shaped distribution with a fixed proportion of observations lying within a certain distance of the mean. This distance is called the **standard deviation (SD)** and reflects the **degree of dispersion** from the mean. According to the properties of this distribution, 68% of observations lie within 1 SD on either side of the mean, with half (ie, 68/2 = 34%) above and half (34%) below the mean. The remaining 32% (= 100% − 68%) lie outside 1 SD from the mean, with half (ie, 32/2 = 16%) above and the other half (16%) below 1 SD from the mean. In addition, 95% of all observations lie within 2 SDs of the mean, and 99.7% of all observations lie within 3 SDs of the mean. This is the **68-95-99.7 rule**.

For men age 35-44 in this sample, the mean systolic blood pressure (SBP) is 120 mm Hg with an SD of 20 mm Hg. Therefore, subjects with SBP >140 mm Hg (to fit the proposed definition of hypertension) represent observations >1 SD above the mean. Based on the 68-95-99.7 rule, 32% of observations (= 100% − 68%) lie outside 1 SD from the mean, with **half (16%) above** and half (16%) below 1 SD from the mean. Therefore, 16% of men age 35-44 have SBP >140 mm Hg.

(Choices B, C, D, and E) Based on the 68-95-99.7 rule, 34% of subjects lie either 1 SD above the mean or 1 SD below the mean. In men age 35-44:

- 34% have SBP 100-120 mm Hg, 34% have SBP 120-140 mm Hg, and, 68% have SBP 100-140 mm Hg.
- Half (50%) have SBP above the mean (ie, >120 mm Hg) and half have SBP below the mean (<120 mm Hg).
- Given that 84% = 50% + 34%, this could represent the sum of those with SBP 100-120 mm Hg (34%) and those with SBP >120 mm Hg (50%): in other words, those with SBP >100 mm Hg.

(Choice F) Based on the 68-95-99.7% rule, 95% of observations lie within 2 SDs (eg, 2 × 20 = 40 mm Hg) of the mean (eg, 120 mm Hg) (ie, SBP between 80 mm Hg [= 120 − 40] and 160 mm Hg [= 120 + 40]).

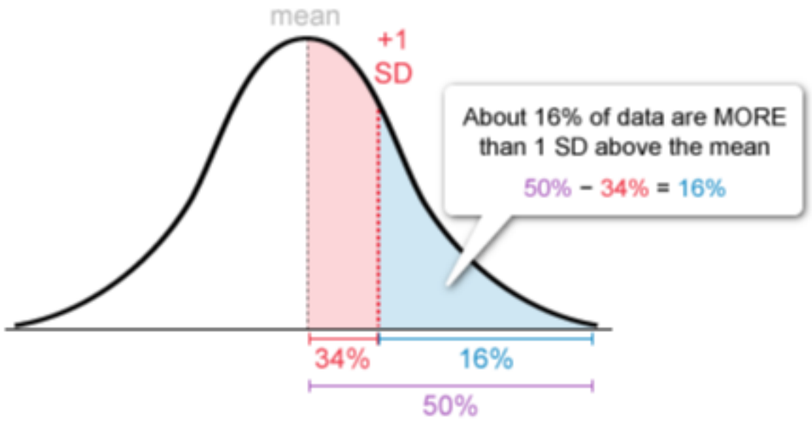
Educational objective:
In a normal (bell-shaped) distribution, 68% of all values are within 1 standard deviation (SD) of the mean; 95% are within 2 SD of the mean; and 99.7% are within 3 SD of the mean.

References

- Our obsession with normal values.

©UWorld

Exhibit Display



Zoom In
Zoom Out
Reset
Add To Flash Card

below 1 SD from the mean. In addition, 95% of all observations lie within 2 SDs of the mean, and 99.7% of all observations lie within 3 SDs of the mean. This is the **68-95-99.7 rule**.

For men age 35-44 in this sample, the mean systolic blood pressure (SBP) is 120 mm Hg with an SD of 20 mm Hg. Therefore, subjects with SBP >140 mm Hg (to fit the proposed definition of hypertension) represent observations >1 SD above the mean. Based on the 68-95-99.7 rule, 32% of observations (= 100% − 68%) lie outside 1 SD from the mean, with **half (16%) above** and half (16%) below 1 SD from the mean. Therefore, 16% of men age 35-44 have SBP >140 mm Hg.

(Choices B, C, D, and E) Based on the 68-95-99.7 rule, 34% of subjects lie either 1 SD above the mean or 1 SD below the mean. In men age 35-44:

- 34% have SBP 100-120 mm Hg, 34% have SBP 120-140 mm Hg, and, 68% have SBP 100-140 mm Hg.
- Half (50%) have SBP above the mean (ie, >120 mm Hg) and half have SBP below the mean (<120 mm Hg).
- Given that 84% = 50% + 34%, this could represent the sum of those with SBP 100-120 mm Hg (34%) and those with SBP >120 mm Hg (50%): in other words, those with SBP >100 mm Hg.

(Choice F) Based on the 68-95-99.7% rule, 95% of observations lie within 2 SDs (eg, 2 × 20 = 40 mm Hg) of the mean (eg, 120 mm Hg) (ie, SBP between 80 mm Hg [= 120 − 40] and 160 mm Hg [= 120 + 40]).

Educational objective:
In a normal (bell-shaped) distribution, 68% of all values are within 1 standard deviation (SD) of the mean; 95% are within 2 SD of the mean; and 99.7% are within 3 SD of the mean.

References

- Our obsession with normal values.

Biostatistics	Biostatistics & Epidemiology	Normal distribution
Subject	System	Topic

Copyright © UWorld. All rights reserved.

A randomized controlled trial is conducted to compare transcatheter aortic valve replacement (TAVR), using a self-expanding transcatheter aortic valve bioprosthesis, with surgical aortic valve replacement in patients with severe aortic stenosis. The objective is to determine the difference in risk of death from any cause. Seven hundred forty-seven patients with severe aortic stenosis who are at increased surgical risk are recruited and randomly assigned to groups for treatment with TAVR or surgical valve replacement. Follow-up assessments are performed at discharge and at 1 month, 6 months, and 1 year. The 1-year follow-up results are as follows:

	Death from any cause		
	Yes	**No**	**Total**
TAVR	57	333	390
Surgical valve replacement	68	289	357
Total	125	622	747

Which of the following best represents the absolute risk reduction for death from any cause in patients treated with TAVR compared to those treated with surgical replacement?

A. (68/357) − (57/390) = 0.044

B. [(68/357) − (57/390)] / (68/357) = 0.233

C. (57/390) / (68/357) = 0.767

D. (68/357) / (57/390) = 1.303

E. 1 / [(68/357) − (57/390)] = 22.562

Submit

A randomized controlled trial is conducted to compare transcatheter aortic valve replacement (TAVR), using a self-expanding transcatheter aortic valve bioprosthesis, with surgical aortic valve replacement in patients with severe aortic stenosis. The objective is to determine the difference in risk of death from any cause. Seven hundred forty-seven patients with severe aortic stenosis who are at increased surgical risk are recruited and randomly assigned to groups for treatment with TAVR or surgical valve replacement. Follow-up assessments are performed at discharge and at 1 month, 6 months, and 1 year. The 1-year follow-up results are as follows:

	Death from any cause		
	Yes	No	Total
TAVR	57	333	390
Surgical valve replacement	68	289	357
Total	125	622	747

Which of the following best represents the absolute risk reduction for death from any cause in patients treated with TAVR compared to those treated with surgical replacement?

- A. (68/357) − (57/390) = 0.044 (68%)
- B. [(68/357) − (57/390)] / (68/357) = 0.233 (10%)
- C. (57/390) / (68/357) = 0.767 (9%)
- D. (68/357) / (57/390) = 1.303 (2%)
- E. 1 / [(68/357) − (57/390)] = 22.562 (8%)

Omitted
Correct answer

68%

04 secs

06/22/2020

Explanation

Absolute risk reduction (ARR) describes the difference in risk of an unfavorable outcome between the treatment group (eg, new treatment) and the control group (eg, standard treatment):

ARR = ($Risk_{control}$ − $Risk_{treatment}$)

In this example, **$Risk_{control}$** is the risk of death from any cause in the group treated with surgical aortic valve replacement and is determined by dividing the number of deaths in that group (68) by the total number of patients in that group (357). **$Risk_{treatment}$** is the risk of death from any cause in the group treated with transcatheter aortic valve replacement (TAVR) and is determined by dividing 57 by 390. Therefore, the ARR for death from any cause in patients treated with TAVR compared to those treated with surgical replacement is calculated as follows:

ARR = (68/357) − (57/390) = 0.044

ARR can be used to determine the number needed to treat (NNT), calculated as:

NNT = 1 / ARR

NNT is the number of patients who need to be treated with a treatment (eg, TAVR) to prevent 1 additional negative event (eg, death from any cause) when compared to a control group (eg, surgical replacement). A lower NNT indicates a more effective treatment because fewer patients need to be treated to prevent 1 additional negative outcome. In this example, the approximate NNT with TAVR to prevent 1 additional death from any cause is:

NNT = 1 / [(68/357) − (57/390)] = 22.562 **(Choice E)**

(Choice B) [(68/357) − (57/390)] / (68/357) = 0.233 represents the relative risk reduction (RRR) for death from any cause in the TAVR group compared to the surgical replacement group: RRR = ($Risk_{control}$ − $Risk_{treatment}$) / $Risk_{control}$. Unlike ARR, which quantifies the absolute difference in risk between the groups, RRR describes the difference in risk between the groups as a proportion of baseline risk (ie, risk in the control group).

(Choices C and D) Switching the numerator and denominator changes the interpretation of the relative risk (RR, the ratio of the probability of an event occuring in one group compared to the other group). In this example, (57/390) / (68/357) represents the RR of death from any cause in the TAVR group compared to the surgical replacement group; (68/357) / (57/390) represents the RR of death from any cause in the surgical replacement group compared to the TAVR group.

NNT = 1 / ARR

NNT is the number of patients who need to be treated with a treatment (eg, TAVR) to prevent 1 additional negative event (eg, death from any cause) when compared to a control group (eg, surgical replacement). A lower NNT indicates a more effective treatment because fewer patients need to be treated to prevent 1 additional negative outcome. In this example, the approximate NNT with TAVR to prevent 1 additional death from any cause is:

NNT = 1 / [(68/357) − (57/390)] = 22.562 **(Choice E)**

(Choice B) [(68/357) − (57/390)] / (68/357) = 0.233 represents the relative risk reduction (RRR) for death from any cause in the TAVR group compared to the surgical replacement group: RRR = ($Risk_{control}$ − $Risk_{treatment}$) / $Risk_{control}$. Unlike ARR, which quantifies the absolute difference in risk between the groups, RRR describes the difference in risk between the groups as a proportion of baseline risk (ie, risk in the control group).

(Choices C and D) Switching the numerator and denominator changes the interpretation of the relative risk (RR, the ratio of the probability of an event occuring in one group compared to the other group). In this example, (57/390) / (68/357) represents the RR of death from any cause in the TAVR group compared to the surgical replacement group; (68/357) / (57/390) represents the RR of death from any cause in the surgical replacement group compared to the TAVR group.

Educational objective:
Absolute risk reduction (ARR) describes the difference in risk between control and treatment groups. It is calculated as follows: ARR = ($Risk_{control}$ − $Risk_{treatment}$).

References

- Measures of effect in epidemiological research.
- Measures of effect: relative risks, odds ratios, risk difference, and "number needed to treat."

Biostatistics	Biostatistics & Epidemiology	Risk, rate, prevalence and incidence
Subject	System	Topic

Copyright © UWorld. All rights reserved.

An investigator is conducting a randomized, double-blind, placebo-controlled clinical trial of a new drug for the treatment of peripheral neuropathy in adults with multiple myeloma (MM). One hundred fifty patients with MM are enrolled in the trial and randomized to receive either the new drug (n = 75) or placebo (n = 75). Trial protocol requires that patients in both groups take 1 pill per day and keep a pain diary. After 3 months of treatment, each patient is interviewed, and the pain diaries are reviewed; 9 patients taking the new drug and 3 patients taking placebo did not take the pills as instructed. The investigator decides to conduct an intention-to-treat analysis of the study data. Which of the following best describes how the data pertaining to all patients who did not adhere to protocol should be treated?

A. Conduct separate analyses of the 12 nonadherent patients and the 138 adherent patients

B. Exclude all 12 nonadherent patients from analysis

C. Exclude the 3 nonadherent patients in the group taking placebo from analysis

D. Exclude the 9 nonadherent patients in the group taking the new drug from analysis

E. Keep all 12 nonadherent patients in their respective groups for analysis

Submit

An investigator is conducting a randomized, double-blind, placebo-controlled clinical trial of a new drug for the treatment of peripheral neuropathy in adults with multiple myeloma (MM). One hundred fifty patients with MM are enrolled in the trial and randomized to receive either the new drug (n = 75) or placebo (n = 75). Trial protocol requires that patients in both groups take 1 pill per day and keep a pain diary. After 3 months of treatment, each patient is interviewed, and the pain diaries are reviewed; 9 patients taking the new drug and 3 patients taking placebo did not take the pills as instructed. The investigator decides to conduct an intention-to-treat analysis of the study data. Which of the following best describes how the data pertaining to all patients who did not adhere to protocol should be treated?

A. Conduct separate analyses of the 12 nonadherent patients and the 138 adherent patients (12%)

B. Exclude all 12 nonadherent patients from analysis (28%)

C. Exclude the 3 nonadherent patients in the group taking placebo from analysis (1%)

D. Exclude the 9 nonadherent patients in the group taking the new drug from analysis (3%)

✓ E. Keep all 12 nonadherent patients in their respective groups for analysis (55%)

Omitted
Correct answer
E

55%
Answered correctly

02 secs
Time Spent

06/20/2020
Last Updated

Explanation

An **intention-to-treat** (ITT) analysis compares treatment groups in a randomized trial by **including all subjects** as initially allocated after randomization **regardless of what happens** during the study period. The rationale is that if subjects have such a poor treatment response (or are so inconvenienced by treatment administration or side effects) as to not follow protocol specifications or drop out of the study, then their outcomes should be attributed to that intervention (ie, nonadherence is related to the practical benefits of treatment). ITT analysis also helps avoid the

Explanation

An **intention-to-treat** (ITT) analysis compares treatment groups in a randomized trial by **including all subjects** as initially allocated after randomization **regardless of what happens** during the study period. The rationale is that if subjects have such a poor treatment response (or are so inconvenienced by treatment administration or side effects) as to not follow protocol specifications or drop out of the study, then their outcomes should be attributed to that intervention (ie, nonadherence is related to the practical benefits of treatment). ITT analysis also helps avoid the effects of attrition (eg, loss to follow-up, dropout) and crossover (eg, switching to another assigned intervention), which may disrupt the benefit of randomization and introduce bias in the estimation of the effect of the intervention.

ITT analysis may lead to a more **conservative estimate** of the effect of the intervention. If attrition or crossover is significant, ITT analysis may be less likely to identify a statistically significant difference between treatments (ie, shift toward the null hypothesis). However, results will reflect the **expected effect** of the intervention in a **practical clinical setting**.

(Choices A, C, and D) In accordance with ITT analysis, the investigator in this case should keep the 12 nonadherent patients in their respective groups for analysis.

(Choice B) An analysis that includes only those patients who strictly adhered to a research protocol (ie, excluding all nonadherent patients from analysis) follows a per-protocol (PP) principle. In PP analysis, the benefit of randomization can be lost with significant subject attrition (eg, sicker subjects selectively dropping out, introducing bias). PP analysis can also overestimate the effect of the intervention in a practical clinical setting, particularly when the treatment is somehow aversive to patients.

Educational objective:
Intention-to-treat analysis includes each subject in their initial randomization group even if subjects stop the intervention or shift to a different intervention. This approach tends to provide a conservative but more valid estimate of the intervention effect in real-world scenarios (ie, clinical settings).

Biostatistics	Biostatistics & Epidemiology	Randomized control trials
Subject	System	Topic

Copyright © UWorld. All rights reserved.

A study is designed to describe the manifestations of coronavirus disease 2019 (COVID-19) on imaging studies, particularly on computerized tomography (CT) scans. Eleven patients with COVID-19 are recruited for the study, and their CT findings are studied and characterized. Which of the following best describes this study design?

A. Case-control
B. Case series
C. Clinical trial
D. Cohort
E. Cross-sectional

Submit

A study is designed to describe the manifestations of coronavirus disease 2019 (COVID-19) on imaging studies, particularly on computerized tomography (CT) scans. Eleven patients with COVID-19 are recruited for the study, and their CT findings are studied and characterized. Which of the following best describes this study design?

- A. Case-control (4%)
- ✓ B. Case series (70%)
- C. Clinical trial (2%)
- D. Cohort (5%)
- E. Cross-sectional (16%)

Omitted
Correct answer
B

70%
Answered correctly

02 secs
Time Spent

06/20/2020
Last Updated

Explanation

Research studies are broadly classified as having an observational design or an experimental design depending on the control the researchers have over the independent variables.

- In **observational studies** (eg, case series, case-control studies, cohort studies, cross-sectional studies), the researchers have **no control** over the independent variables (eg, exposure to risk factors, treatments).
- In experimental studies (eg, crossover design, randomized controlled trials), the researchers control and randomly assign the independent variables (eg, exposure to risk factors, treatments).

Research studies are broadly classified as having an observational design or an experimental design depending on the control the researchers have over the independent variables.

- In **observational studies** (eg, case series, case-control studies, cohort studies, cross-sectional studies), the researchers have **no control** over the independent variables (eg, exposure to risk factors, treatments).
- In experimental studies (eg, crossover design, randomized controlled trials), the researchers control and randomly assign the independent variables (eg, exposure to risk factors, treatments).

Case series is a descriptive observational study design in which a (generally small) group of patients with a similar diagnosis or treatment is described at a point in time or followed over a certain time period. Contrary to other observational designs (eg, cohort, case-control), a case series has **no comparison group (Choices A and D)**. For this reason, a case series cannot establish associations between risk factors (eg, treatments) and outcomes (eg, diseases).

In this example, the CT scan findings of 11 patients with coronavirus disease 2019 (COVID-19) are being studied. The study has no control group; therefore, it can only describe COVID-19 manifestations on imaging studies.

(Choice C) A clinical trial is an experimental study in which patients are prospectively assigned to ≥2 interventions (often including a placebo or control treatment) to evaluate the effects of those interventions on outcomes of interest.

(Choice E) A cross-sectional study is an observational study that assesses a population of interest (eg, all women of reproductive age) at a single point in time; it can be used to estimate disease prevalence and association with risk factors of interest. However, it is not used to characterize disease findings because subjects are not selected based on the presence of disease.

Educational objective:
A case series is a descriptive observational study design in which a group of patients with a similar diagnosis or treatment is described at a point in time or followed over a certain period. This study design has no comparison group; therefore, it cannot establish associations between risk factors (eg, treatments) and outcomes (eg, diseases).

Biostatistics Biostatistics & Epidemiology Study designs

A study investigating a new test for diagnosing acute myocardial infarction (AMI) has just been initiated. The sensitivity of the test is estimated at 75% and the specificity at 80%. The study enrolls 600 patients, of whom 200 are confirmed AMI cases as determined by the diagnostic gold standard. How many false negatives are to be expected in the study?

A. 50

B. 80

C. 120

D. 150

E. 400

Submit

A study investigating a new test for diagnosing acute myocardial infarction (AMI) has just been initiated. The sensitivity of the test is estimated at 75% and the specificity at 80%. The study enrolls 600 patients, of whom 200 are confirmed AMI cases as determined by the diagnostic gold standard. How many false negatives are to be expected in the study?

- ✓ A. 50 (53%)
- B. 80 (20%)
- C. 120 (11%)
- D. 150 (11%)
- E. 400 (3%)

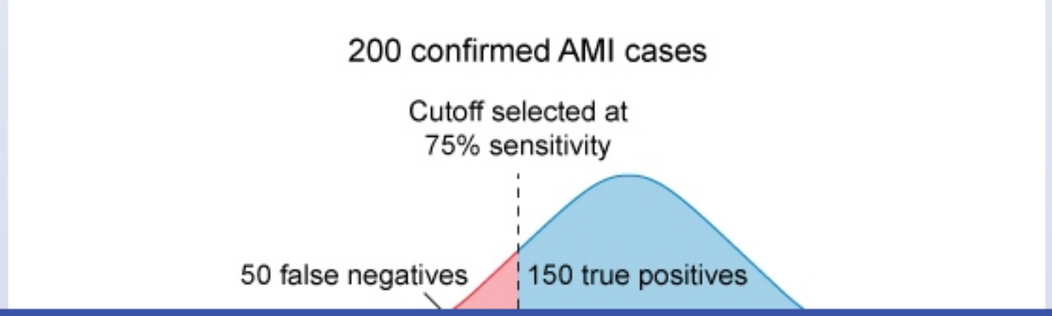
Omitted
Correct answer
A

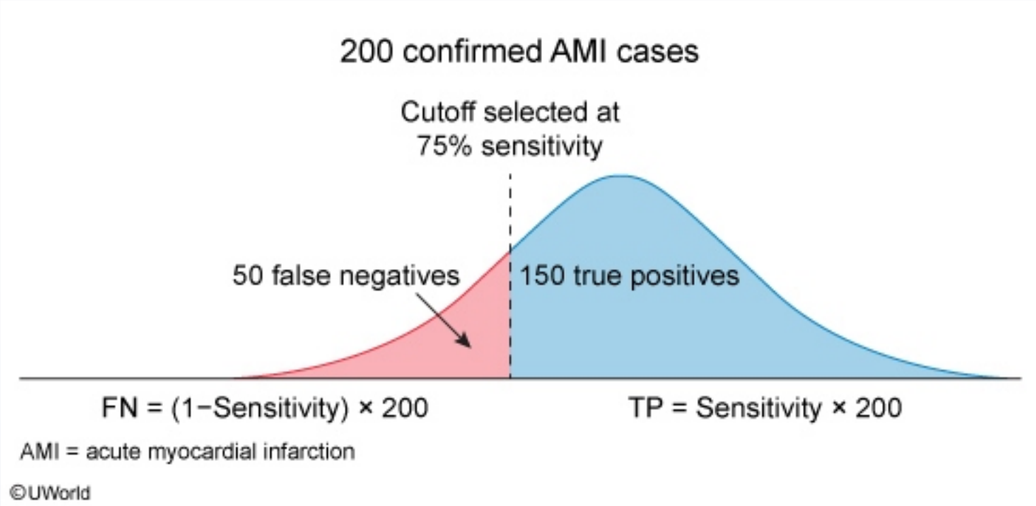
53%
Answered correctly

03 secs
Time Spent

03/15/2020
Last Updated

Explanation





The **sensitivity** of a test determines how well the test categorizes patients who **have the disease** of interest. Tests with higher sensitivity have an increased probability of assigning a positive test result to those patients who truly have the disease; these are considered true positives (rightmost part of the curve). Conversely, patients who **test negative** even though they have the disease are considered **false negatives** (leftmost part of the curve).

True positives (TP) and false negatives (FN) are complementary (ie, an increase in one parameter is exactly offset by a decrease in the other, such that the total is always equal to the number of confirmed patients). As such, the test sensitivity and number of patients with the disease can be used to calculate the number of each:

TP = (Sensitivity) × (Number of patients with the disease)
FN = (1 − Sensitivity) × (Number of patients with the disease)

In this case, the sensitivity of the test is 75% and the total number of patients with confirmed acute myocardial infarction is 200. Therefore, a test that is 75% sensitive will correctly classify 75% of the patients as positive (TP) and incorrectly classify 25% of the patients (1 − Sensitivity) as negative (FN).

The **sensitivity** of a test determines how well the test categorizes patients who **have the disease** of interest. Tests with higher sensitivity have an increased probability of assigning a positive test result to those patients who truly have the disease; these are considered true positives (rightmost part of the curve). Conversely, patients who **test negative** even though they have the disease are considered **false negatives** (leftmost part of the curve).

True positives (TP) and false negatives (FN) are complementary (ie, an increase in one parameter is exactly offset by a decrease in the other, such that the total is always equal to the number of confirmed patients). As such, the test sensitivity and number of patients with the disease can be used to calculate the number of each:

TP = (Sensitivity) × (Number of patients with the disease)
FN = (1 − Sensitivity) × (Number of patients with the disease)

In this case, the sensitivity of the test is 75% and the total number of patients with confirmed acute myocardial infarction is 200. Therefore, a test that is 75% sensitive will correctly classify 75% of the patients as positive (TP) and incorrectly classify 25% of the patients (1 − Sensitivity) as negative (FN).

Using this information, the number of FN can be calculated as follows (alternate method):

FN = (1 − Sensitivity) × (Number of patients with the disease) = (1 − 0.75) × 200 = **50**

Educational objective:
When undergoing diagnostic testing, patients with the disease can test positive (true positive, TP) or negative (false negative, FN). The sensitivity of a test determines the proportion of patients that are correctly classified:

TP = (Sensitivity) × (Number of patients with the disease)
FN = (1 − Sensitivity) × (Number of patients with the disease)

Biostatistics	Cardiovascular System	Sensitivity and specificity
Subject	System	Topic

Copyright © UWorld. All rights reserved.

The **sensitivity** of a test determines how well the test categorizes patients who **have the disease** of interest. Tests with higher sensitivity have an

Exhibit Display

False negatives describe people who actually have a disease but test negative.

	Disease positive	Disease negative	
Test positive	a True positive	b False positive	a + b
Test negative	c False negative	d True negative	c + d
	a + c	b + d	

Sensitivity = a/(a + c)
Specificity = d/(b + d)

The number of false negatives can be calculated using the test's sensitivity and total number of patients with acute myocardial infarction. The sensitivity of a test refers to the number of true positives divided by the total number of patients with the disease (a / [a + c]). In this case, the sensitivity is 75% and the total number of patients with AMI (a + c) equals 200. Using this information, the number of true positives (a) can be calculated as follows:

Sensitivity = a/(a + c)
0.75 = a/200
a = 150

Because the total number of patients with the disease (a + c) equals 200 and the number of true positives (a) equals 150, the number of false negatives (c) is 50.

©UWorld

Zoom In | Zoom Out | Reset | Add To New Card | Existing Card

A 50-year-old man comes to the physician for a routine check-up. He has a 20-pack-year smoking history and consumes alcohol occasionally. He eats a balanced diet and consumes fried foods and red meat several times a week. He is not very active and spends much of his free time watching television and sitting at his computer. His family history is significant for hypertension in his mother and bladder cancer in his father. His blood pressure is 160/90 mm Hg and heart rate is 70/min. Laboratory testing is significant for a fasting blood glucose level of 155 mg/dL. Which of the following interventions is likely to have the greatest effect on reducing this patient's mortality risk?

A. Annual fecal occult blood test

B. Annual prostate-specific antigen test

C. Daily aspirin

D. Dietary modification

E. Regular exercise

F. Smoking cessation

G. Tight glycemic control

Submit

A 50-year-old man comes to the physician for a routine check-up. He has a 20-pack-year smoking history and consumes alcohol occasionally. He eats a balanced diet and consumes fried foods and red meat several times a week. He is not very active and spends much of his free time watching television and sitting at his computer. His family history is significant for hypertension in his mother and bladder cancer in his father. His blood pressure is 160/90 mm Hg and heart rate is 70/min. Laboratory testing is significant for a fasting blood glucose level of 155 mg/dL. Which of the following interventions is likely to have the greatest effect on reducing this patient's mortality risk?

A. Annual fecal occult blood test (0%)

B. Annual prostate-specific antigen test (0%)

C. Daily aspirin (3%)

D. Dietary modification (7%)

E. Regular exercise (11%)

✓ F. Smoking cessation (67%)

G. Tight glycemic control (8%)

Omitted
Correct answer
F

67%
Answered correctly

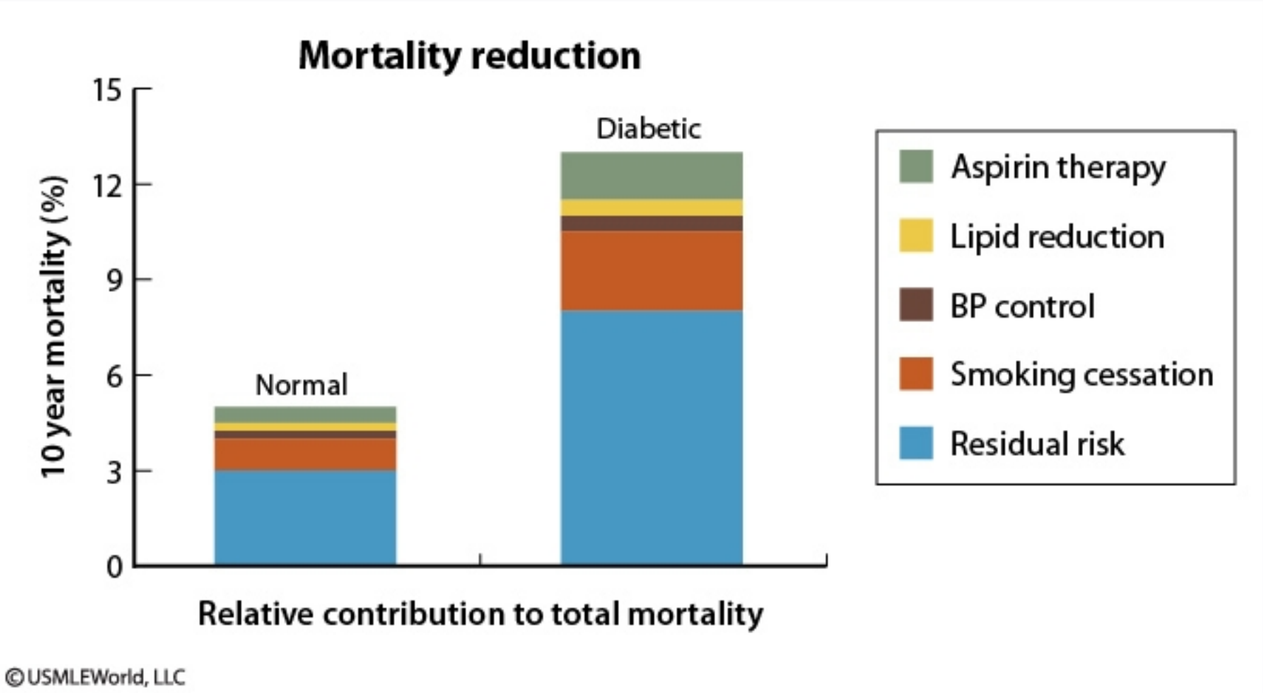
02 secs
Time Spent

03/07/2020
Last Updated

Explanation

Mortality reduction

15



One of the objectives for the national Healthy People 2020 initiative is to reduce the prevalence of tobacco smoking in adults from 20% in 2008 to less than 12% by 2020. Tobacco use, especially smoking, is the single most preventable cause of death and disease in the United States. In addition to its numerous direct adverse effects, tobacco smoking also substantially increases the risk for macrovascular (eg, myocardial infarction [MI], stroke) and microvascular (eg, retinopathy, nephropathy) complications of diabetes mellitus. These complications significantly increase mortality, primarily by increasing the risk of MI. The risk of MI-associated mortality begins to decrease immediately upon smoking cessation. However, it takes several tobacco-free years before the risk returns to baseline, especially in patients with a long history of smoking. In addition, smokers with diabetes have a higher likelihood of developing hypercholesterolemia, poor glycemic control, and end-stage renal disease. It appears that smoking not only worsens the complications of diabetes but also increases the likelihood of developing diabetes in the first place, thus highlighting the importance of prevention and early cessation.

(Choice A) Fecal occult blood testing screens for colorectal cancer and improves mortality through earlier detection. Even so, the mortality

One of the objectives for the national Healthy People 2020 initiative is to reduce the prevalence of tobacco smoking in adults from 20% in 2008 to less than 12% by 2020. Tobacco use, especially smoking, is the single most preventable cause of death and disease in the United States. In addition to its numerous direct adverse effects, tobacco smoking also substantially increases the risk for macrovascular (eg, myocardial infarction [MI], stroke) and microvascular (eg, retinopathy, nephropathy) complications of diabetes mellitus. These complications significantly increase mortality, primarily by increasing the risk of MI. The risk of MI-associated mortality begins to decrease immediately upon smoking cessation. However, it takes several tobacco-free years before the risk returns to baseline, especially in patients with a long history of smoking. In addition, smokers with diabetes have a higher likelihood of developing hypercholesterolemia, poor glycemic control, and end-stage renal disease. It appears that smoking not only worsens the complications of diabetes but also increases the likelihood of developing diabetes in the first place, thus highlighting the importance of prevention and early cessation.

(Choice A) Fecal occult blood testing screens for colorectal cancer and improves mortality through earlier detection. Even so, the mortality benefit is much lower than that achieved by smoking cessation.

(Choice B) Prostate-specific antigen is increased in some patients with prostate cancer. However, annual prostate-specific antigen testing has not been shown to improve mortality from prostate cancer.

(Choice C) Aspirin use significantly reduces the risk of death from coronary heart disease in diabetic patients, at levels similar to smoking cessation. However, when all-cause mortality is compared, smoking cessation has a greater effect on reducing mortality than daily aspirin therapy.

(Choices D, E, and G) Dietary modification and exercise reduce the risk of cardiovascular disease (to a lesser extent than smoking cessation) and help improve glycemic control in diabetic individuals. Although tight glycemic control has been shown to reduce the risk of microvascular complications from diabetes, it does not significantly improve cardiovascular or all-cause mortality.

Educational objective:
Smoking cessation is by far the most effective preventive intervention in almost all patients, and this is especially true in those with diabetes.

References

- How can we best prolong life? Benefits of coronary risk factor reduction in non-diabetic and diabetic subjects.
- Effect of tight blood glucose control versus conventional control in patients with type 2 diabetes mellitus: a systematic review with meta-analysis of randomized controlled trials.

400 women aged 20-35 coming for routine check-up are asked about their smoking status. 40% of the women are smokers. Over the next ten years, 25 smokers and 24 non-smokers developed breast cancer. Which of the following best describes the study design?

A. Prospective cohort study

B. Retrospective cohort study

C. Case-control study

D. Cross-sectional study

E. Randomized clinical trial

Submit

400 women aged 20-35 coming for routine check-up are asked about their smoking status. 40% of the women are smokers. Over the next ten years, 25 smokers and 24 non-smokers developed breast cancer. Which of the following best describes the study design?

- ✓ A. Prospective cohort study (85%)
- B. Retrospective cohort study (4%)
- C. Case-control study (6%)
- D. Cross-sectional study (3%)
- E. Randomized clinical trial (0%)

Omitted
Correct answer
A

85%
Answered correctly

02 secs
Time Spent

02/21/2020
Last Updated

Explanation

The scenario described above is a good example of prospective cohort study. Initially a group of subjects is selected (i.e., cohort) and their exposure status is determined (smoker/non-smoker). The cohort is then followed for a certain period of time and observed for development of the outcome (breast cancer). Sometimes, the exposure status is determined retrospectively and then patients are tracked from that point of time, typically using medical records **(Choice B)**.

(Choice C) A case-control study is designed by selecting patients with a particular disease (cases) and without that disease (controls) and then determining their previous exposure status.

(Choice D) A cross-sectional study is also known as a prevalence study. It is characterized by the simultaneous measurement of exposure and

Omitted
Correct answer
A

85%
Answered correctly

02 secs
Time Spent

02/21/2020
Last Updated

Explanation

The scenario described above is a good example of prospective cohort study. Initially a group of subjects is selected (i.e., cohort) and their exposure status is determined (smoker/non-smoker). The cohort is then followed for a certain period of time and observed for development of the outcome (breast cancer). Sometimes, the exposure status is determined retrospectively and then patients are tracked from that point of time, typically using medical records **(Choice B)**.

(Choice C) A case-control study is designed by selecting patients with a particular disease (cases) and without that disease (controls) and then determining their previous exposure status.

(Choice D) A cross-sectional study is also known as a prevalence study. It is characterized by the simultaneous measurement of exposure and outcome. It is a snapshot study design that frequently uses surveys. They are relatively inexpensive and easy to perform.

(Choice E) A randomized clinical trial directly compares two or more treatments. Usually, the subjects are randomly assigned to an exposure (e.g., a medication) or placebo and then followed for the development of the outcome of interest.

Educational Objective:
Prospective cohort studies are organized by selecting a group of individuals (i.e., cohort), determining their exposure status, and then following them over time for development of the disease of interest.

Biostatistics	Hematology & Oncology	Study designs
Subject	System	Topic

Copyright © UWorld. All rights reserved.

Plasma homocysteine levels are measured in patients with acute coronary syndrome who are treated at a large community hospital. The mean plasma homocysteine level in this group is determined to be 11.1 μmol/L with a standard deviation of 1.2 μmol/L. In a separate group of patients hospitalized on the general ward in the same hospital, the mean plasma level is 9.5 μmol/L and the standard deviation is 1.3 μmol/L. Which of the following statistical methods should be used to compare the mean homocysteine levels of these 2 groups of patients?

A. Two-sample t test

B. Linear regression

C. Correlation coefficient

D. Chi-square test

E. Meta-analysis

Submit

Plasma homocysteine levels are measured in patients with acute coronary syndrome who are treated at a large community hospital. The mean plasma homocysteine level in this group is determined to be 11.1 μmol/L with a standard deviation of 1.2 μmol/L. In a separate group of patients hospitalized on the general ward in the same hospital, the mean plasma level is 9.5 μmol/L and the standard deviation is 1.3 μmol/L. Which of the following statistical methods should be used to compare the mean homocysteine levels of these 2 groups of patients?

- ✓ A. Two-sample t test (76%)
- B. Linear regression (1%)
- C. Correlation coefficient (4%)
- D. Chi-square test (16%)
- E. Meta-analysis (1%)

Omitted
Correct answer
A

76%
Answered correctly

03 secs
Time Spent

03/23/2020
Last Updated

Explanation

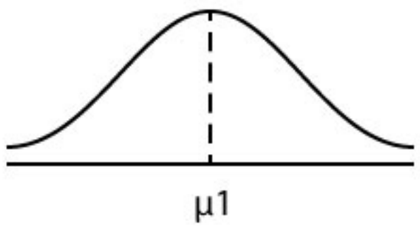
Two-sample t test

Null hypothesis	Alternate hypothesis
Both samples drawn from the same population	Both samples drawn from different populations

Two-sample t test

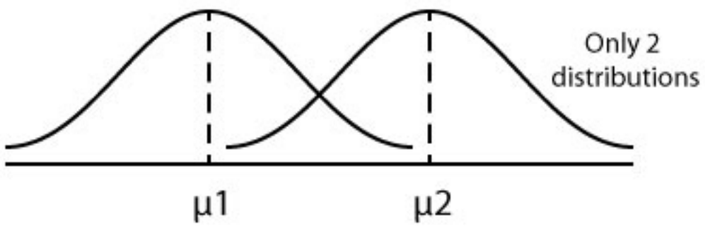
Null hypothesis

Both samples drawn from the same population
($\mu 1 = \mu 2$)

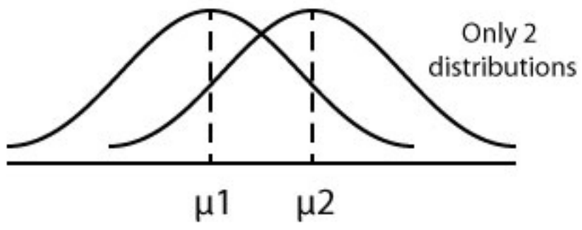


Alternate hypothesis

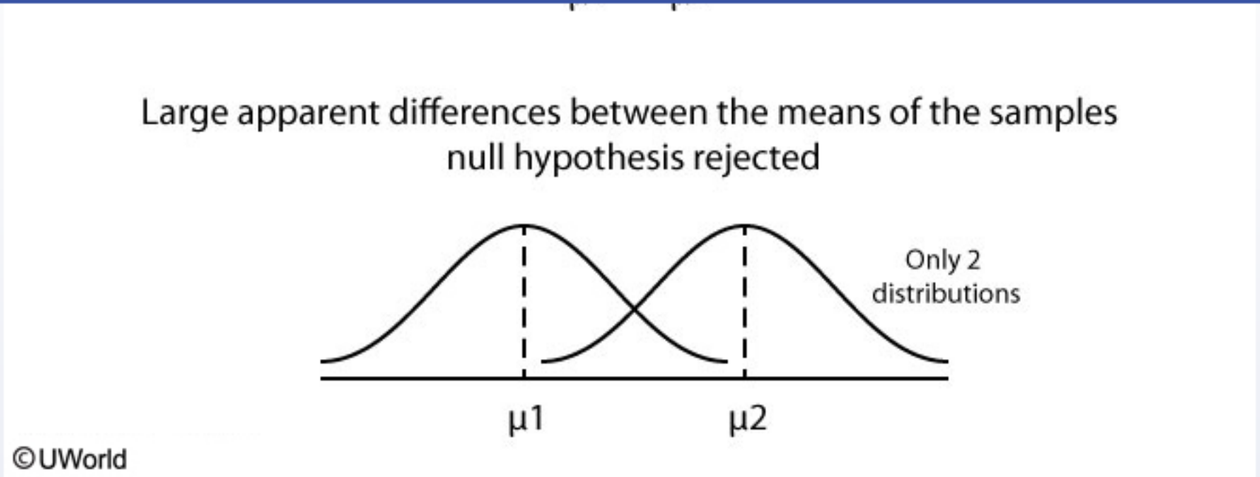
Both samples drawn from different populations
($\mu 1 \neq \mu 2$)



Small apparent difference between the means of the samples
null hypothesis not rejected



Large apparent differences between the means of the samples
null hypothesis rejected



The two-sample t test is commonly employed to determine if the means of 2 populations are equal. Several statistical approaches can be used, but the basic numerical requirements needed to perform this test are the 2 mean values, the sample variances (eg, standard deviations), and the sample sizes. The t statistic is then calculated, from which the p value can be determined. If $p < 0.05$, the null hypothesis (which assumes that there is no difference between 2 groups) is rejected and the 2 means are assumed to be statistically different.

(Choice B) Linear regression is used to model the linear relationship between a dependent variable and an independent variable. For example, linear regression could be used to determine the relationship (described in terms of a trend line) between the number of cigarettes smoked per day and the number of yearly hospitalizations in COPD patients.

(Choice C) The correlation coefficient is a measure of the strength and direction of a linear relationship between 2 variables. For example, a study may report a correlation coefficient describing the association between estrogen levels and breast cancer risk in postmenopausal women. It is different from linear regression in that a single number is reported describing the strength and magnitude of the association.

(Choice D) The chi-square test is most appropriate for use with categorical data. It can be used to evaluate whether the expected frequency of an occurrence is consistent with the observed frequency of that occurrence ("goodness of fit"). For instance, a study evaluating Mendelian inheritance of red and green seed colors would use a chi-square test to compare the observed and expected proportions of each seed type.

(Choice E) Meta-analysis is an epidemiologic method of analyzing pooled data from several studies, thereby increasing the statistical power beyond that of the individual studies.

©UWorld

The two-sample t test is commonly employed to determine if the means of 2 populations are equal. Several statistical approaches can be used, but the basic numerical requirements needed to perform this test are the 2 mean values, the sample variances (eg, standard deviations), and the sample sizes. The t statistic is then calculated, from which the p value can be determined. If $p < 0.05$, the null hypothesis (which assumes that there is no difference between 2 groups) is rejected and the 2 means are assumed to be statistically different.

(Choice B) Linear regression is used to model the linear relationship between a dependent variable and an independent variable. For example, linear regression could be used to determine the relationship (described in terms of a trend line) between the number of cigarettes smoked per day and the number of yearly hospitalizations in COPD patients.

(Choice C) The correlation coefficient is a measure of the strength and direction of a linear relationship between 2 variables. For example, a study may report a correlation coefficient describing the association between estrogen levels and breast cancer risk in postmenopausal women. It is different from linear regression in that a single number is reported describing the strength and magnitude of the association.

(Choice D) The chi-square test is most appropriate for use with categorical data. It can be used to evaluate whether the expected frequency of an occurrence is consistent with the observed frequency of that occurrence ("goodness of fit"). For instance, a study evaluating Mendelian inheritance of red and green seed colors would use a chi-square test to compare the observed and expected proportions of each seed type.

(Choice E) Meta-analysis is an epidemiologic method of analyzing pooled data from several studies, thereby increasing the statistical power beyond that of the individual studies.

Educational objective:
The two-sample t test is a statistical method commonly employed to compare the means of 2 groups of subjects.

Biostatistics	Cardiovascular System	Statistical tests
Subject	System	Topic

Copyright © UWorld. All rights reserved.

A study to assess spironolactone's efficacy in patients with heart failure is performed. 450 patients receive either spironolactone or placebo for two years. Neither the patients nor physicians are aware of who takes the drug or placebo. The study setup described above is most effective in preventing:

A. Beta error

B. Recall bias

C. Observer bias

D. Effect modification

E. Selection bias

Submit

A study to assess spironolactone's efficacy in patients with heart failure is performed. 450 patients receive either spironolactone or placebo for two years. Neither the patients nor physicians are aware of who takes the drug or placebo. The study setup described above is most effective in preventing:

- A. Beta error (2%)
- B. Recall bias (2%)
- ✓ C. Observer bias (75%)
- D. Effect modification (7%)
- E. Selection bias (12%)

Omitted
Correct answer
C

75%
Answered correctly

02 secs
Time Spent

02/03/2020
Last Updated

Explanation

Blinding technique is commonly used in clinical trials. The blinding can involve patients exclusively or both patients and physicians (double blinding). The main purpose of blinding is to prevent patient or researcher expectancy from interfering with the determination of an outcome. For example, a researcher's belief in a positive outcome in treated patients can potentially result in observer bias.

(Choice A) Beta error refers to a conclusion that there is no difference between the groups studied when a difference truly exists. Beta error is a random error, not a systematic error (i.e., bias).

(Choice B) Recall bias results from the inaccurate recall of past exposure by subjects. It applies mostly to case-control studies.

E. Selection bias (12%)

Omitted
Correct answer
C

75%
Answered correctly

02 secs
Time Spent

02/03/2020
Last Updated

Explanation

Blinding technique is commonly used in clinical trials. The blinding can involve patients exclusively or both patients and physicians (double blinding). The main purpose of blinding is to prevent patient or researcher expectancy from interfering with the determination of an outcome. For example, a researcher's belief in a positive outcome in treated patients can potentially result in observer bias.

(Choice A) Beta error refers to a conclusion that there is no difference between the groups studied when a difference truly exists. Beta error is a random error, not a systematic error (i.e., bias).

(Choice B) Recall bias results from the inaccurate recall of past exposure by subjects. It applies mostly to case-control studies.

(Choice D) Effect modification is not a bias and should not be controlled.

(Choice E) Selection bias results from the manner in which people are selected for the study, or from the selective losses from follow-up.

Educational Objective:
The main purpose of blinding is to prevent patient or researcher expectancy from interfering with an outcome.

Biostatistics	Cardiovascular System	Bias
Subject	System	Topic

Copyright © UWorld. All rights reserved.

A 45-year-old woman comes to the office with a palpable thyroid node. Medical history reveals that she was diagnosed with medulloblastoma during childhood and treated with chemotherapy and external beam radiation. Fine-needle aspiration of the thyroid shows no malignant cells. As the test results are explained to the patient, she asks, "What are the chances that I really do not have thyroid cancer?" Which of the following diagnostic test parameters would be most useful for answering this patient's question?

A. Negative predictive value

B. Positive predictive value

C. Sensitivity

D. Specificity

E. Validity

Submit

A 45-year-old woman comes to the office with a palpable thyroid node. Medical history reveals that she was diagnosed with medulloblastoma during childhood and treated with chemotherapy and external beam radiation. Fine-needle aspiration of the thyroid shows no malignant cells. As the test results are explained to the patient, she asks, "What are the chances that I really do not have thyroid cancer?" Which of the following diagnostic test parameters would be most useful for answering this patient's question?

- ✓ A. Negative predictive value (74%)
- B. Positive predictive value (2%)
- C. Sensitivity (9%)
- D. Specificity (11%)
- E. Validity (0%)

Omitted
Correct answer
A

74%
Answered correctly

02 secs
Time Spent

05/05/2020
Last Updated

Explanation

Negative predictive value (NPV) is defined as the probability of not having a disease when the test result is negative. NPV is calculated as the proportion of **true negatives** divided by the total number of **negative tests** (true and false negatives); therefore, it varies with the prevalence of disease in the target population.

The prevalence of a disease in a population may be used as an estimate for the **pretest probability** of having the disease in patients who **closely resemble** the population. A patient with a high pretest probability will have a low NPV with a negative test, whereas a patient with a low pretest probability will have a high NPV with a negative test.

Negative predictive value (NPV) is defined as the probability of not having a disease when the test result is negative. NPV is calculated as the proportion of **true negatives** divided by the total number of **negative tests** (true and false negatives); therefore, it varies with the prevalence of disease in the target population.

The prevalence of a disease in a population may be used as an estimate for the **pretest probability** of having the disease in patients who **closely resemble** the population. A patient with a high pretest probability will have a low NPV with a negative test, whereas a patient with a low pretest probability will have a high NPV with a negative test.

This patient has a high pretest probability for having thyroid cancer since she most likely received significant radiation exposure to the thyroid during childhood. As such, her negative cytology results would have a lower NPV when compared to another patient with no risk factors for thyroid malignancy who also tests negative.

(Choice B) Positive predictive value also varies according to the pretest probability but applies to positive test results.

(Choices C and D) Sensitivity and specificity of a test are intrinsic characteristics of a test; their values are fixed and do not vary with the pretest probability of a disease. An ideal diagnostic test should be both highly sensitive and highly specific. In this case, fine-needle aspiration has a high sensitivity and specificity; however, the NPV is a better parameter for addressing the patient's question (ie, how likely is she to not have thyroid cancer given her negative test result?); in this case her NPV might be low given her prior radiation exposure.

(Choice E) Validity represents the accuracy of the test (ie, the test measures what it is supposed to measure). It does not depend on the pretest probability of the disease.

Educational objective:
Negative predictive value (NPV) is the probability of not having a disease when the test result is negative. The NPV will vary with the pretest probability of a disease. A patient with a high probability of having a disease will have a low NPV with a negative test, but a patient with a low probability of having a disease will have a high NPV with a negative test.

Biostatistics	General Principles	Predictive values
Subject	System	Topic

Copyright © UWorld. All rights reserved.

Negative predictive value (NPV) is defined as the probability of not having a disease when the test result is negative. NPV is calculated as the proportion of **true negatives** divided by the total number of **negative tests** (true and false negatives); therefore, it varies with the prevalence of disease in the target population.

The prevalence of a disease in a population may be used as an estimate for the **pretest probability** of having the disease in patients who **closely resemble** the population. A patient with a high pretest probability will have a low NPV with a negative test, whereas a patient with a low pretest probability will have a high NPV with a negative test.

This patient has a high pretest probability for having thyroid cancer since she most likely received significant radiation exposure to the thyroid during childhood. As such, her negative cytology results would have a lower NPV when compared to another patient with no risk factors for thyroid malignancy who also tests negative.

(Choice B) Positive predictive value also varies according to the pretest probability but applies to positive test results.

(Choices C and D) Sensitivity and specificity of a test are intrinsic characteristics of a test; their values are fixed and do not vary with the pretest probability of a disease. An ideal diagnostic test should be both highly sensitive and highly specific. In this case, fine-needle aspiration has a high sensitivity and specificity; however, the NPV is a better parameter for addressing the patient's question (ie, how likely is she to not have thyroid cancer given her negative test result?); in this case her NPV might be low given her prior radiation exposure.

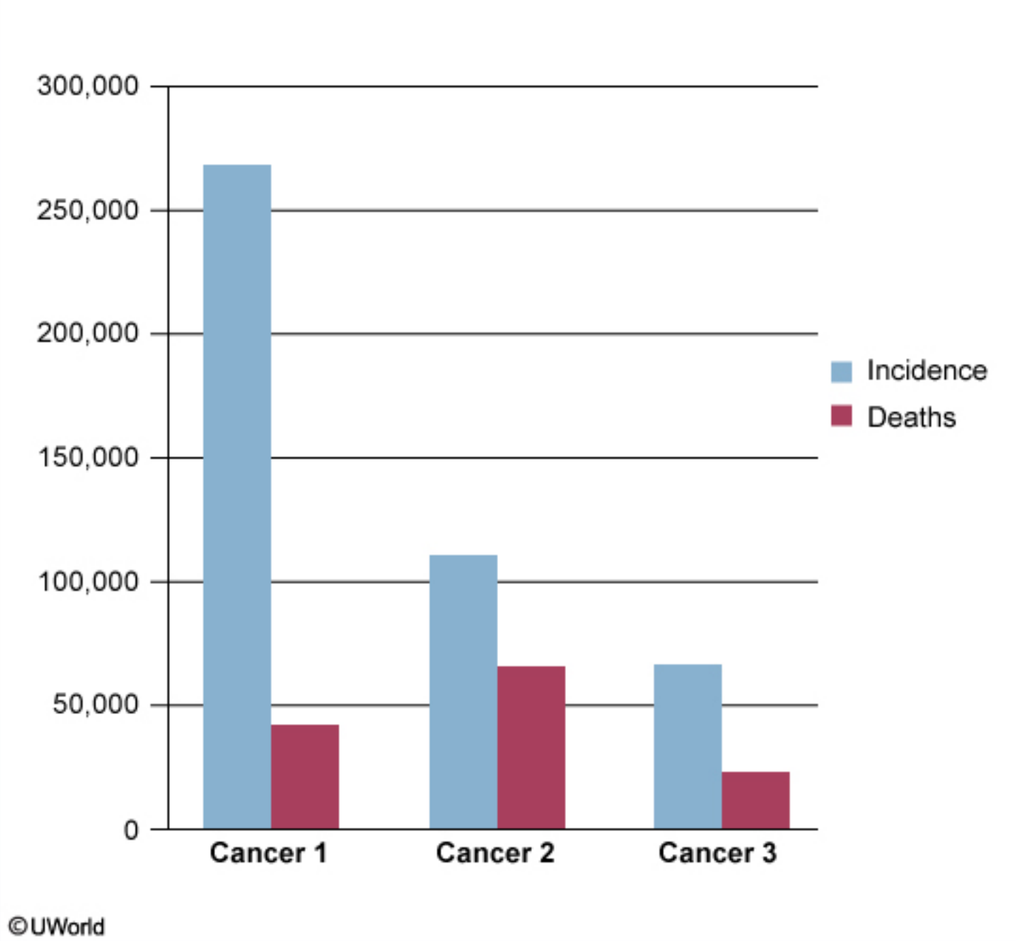
(Choice E) Validity represents the accuracy of the test (ie, the test measures what it is supposed to measure). It does not depend on the pretest probability of the disease.

Educational objective:
Negative predictive value (NPV) is the probability of not having a disease when the test result is negative. The NPV will vary with the pretest probability of a disease. A patient with a high probability of having a disease will have a low NPV with a negative test, but a patient with a low probability of having a disease will have a high NPV with a negative test.

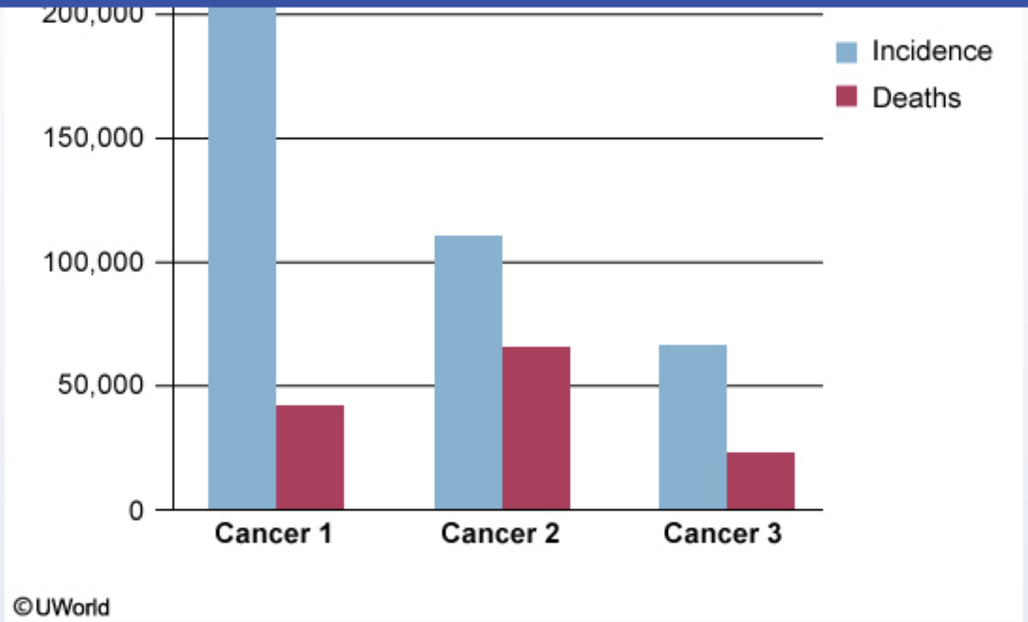
Biostatistics	General Principles	Predictive values
Subject	System	Topic

Copyright © UWorld. All rights reserved.

The figure below shows the estimated incidence and mortality in 2019 for the 3 most common cancers affecting women in the United States, based on the American Cancer Society's *Cancer Facts & Figures 2019*. The x-axis lists the specific cancers, and the y-axis shows the projected number of patients affected.



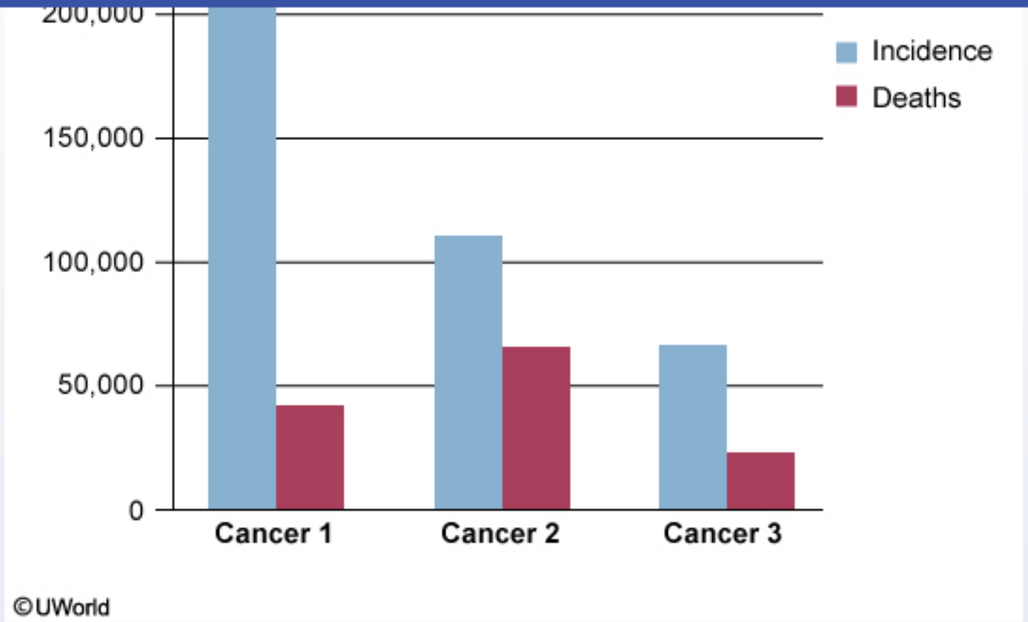
Which of the following lists of cancers most likely corresponds to Cancers 1, 2, and 3, respectively?



Which of the following lists of cancers most likely corresponds to Cancers 1, 2, and 3, respectively?

- A. Breast, lung, colorectal
- B. Lung, breast, colorectal
- C. Breast, ovary, lung
- D. Lung, breast, cervix
- E. Breast, colorectal, lung

Submit



Which of the following lists of cancers most likely corresponds to Cancers 1, 2, and 3, respectively?

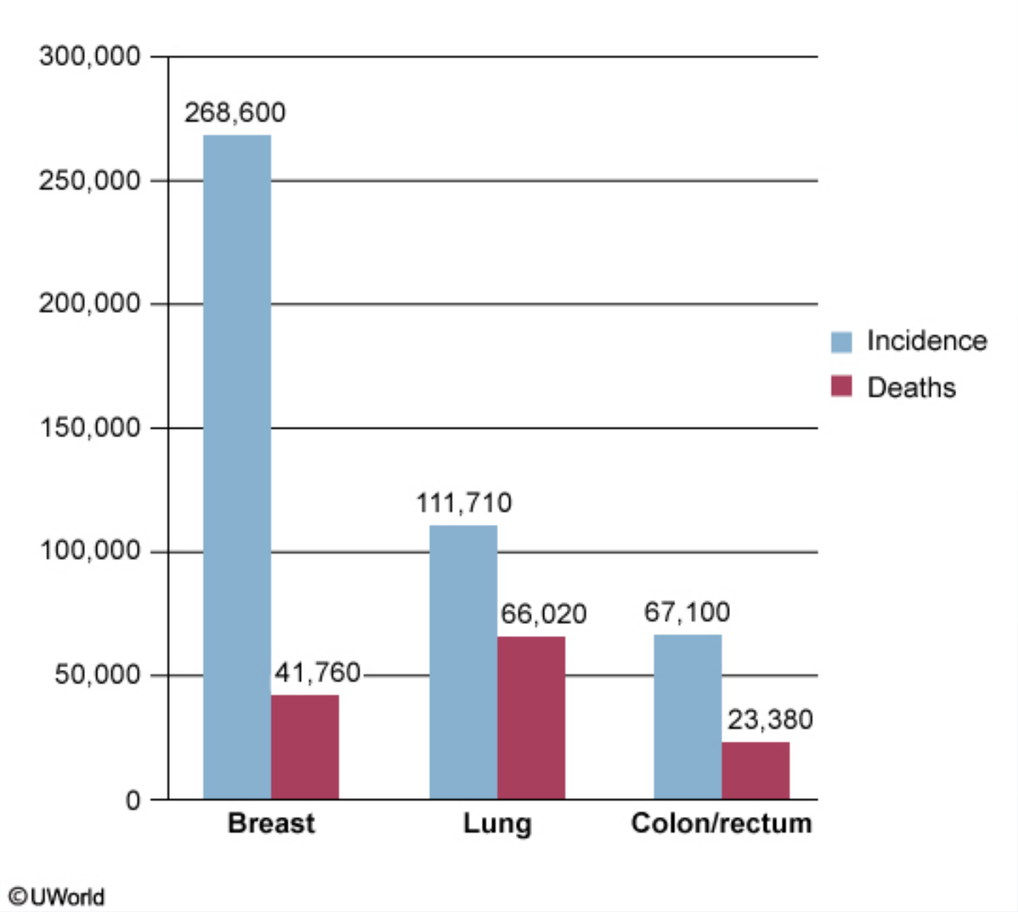
- A. Breast, lung, colorectal (77%)
- B. Lung, breast, colorectal (10%)
- C. Breast, ovary, lung (4%)
- D. Lung, breast, cervix (3%)
- E. Breast, colorectal, lung (4%)

Omitted
Correct answer

77%
Answered correctly

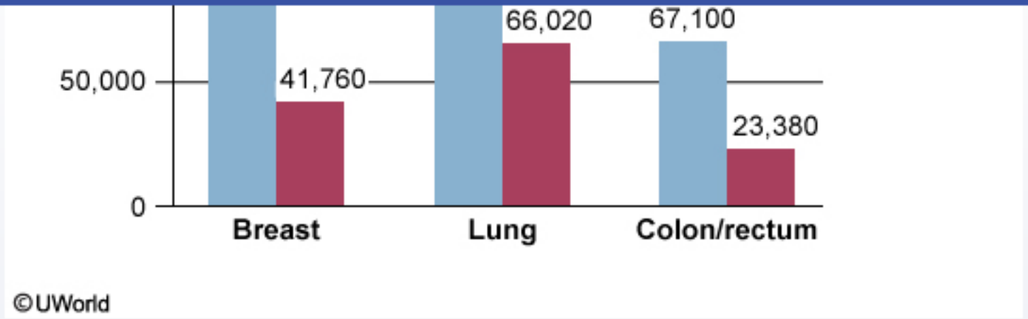
06 secs
Time Spent

02/26/2020
Last Updated



According to estimates by the American Cancer Society, the most common cancers (excluding basal and squamous cell skin cancer) in 2019 among women living in the United States are (by order of incidence) breast, lung, and colorectal cancer. Although **breast cancer** has the **highest incidence** (268,600), its mortality is relatively low (41,760) due to early detection while the tumor is still localized. In contrast, the incidence of **lung cancer** in women is lower (111,710), but its **mortality is higher** (66,020), in part because lung cancer tends to be detected later and is less localized **(Choice B)**. Lung cancer has emerged as the leading cause of cancer mortality in women.

(Choices C and D) The incidence of ovarian (22,530) and cervical (13,170) cancers is much lower than that of the three cancers presented in the



According to estimates by the American Cancer Society, the most common cancers (excluding basal and squamous cell skin cancer) in 2019 among women living in the United States are (by order of incidence) breast, lung, and colorectal cancer. Although **breast cancer** has the **highest incidence** (268,600), its mortality is relatively low (41,760) due to early detection while the tumor is still localized. In contrast, the incidence of **lung cancer** in women is lower (111,710), but its **mortality is higher** (66,020), in part because lung cancer tends to be detected later and is less localized **(Choice B)**. Lung cancer has emerged as the leading cause of cancer mortality in women.

(Choices C and D) The incidence of ovarian (22,530) and cervical (13,170) cancers is much lower than that of the three cancers presented in the figure.

(Choice E) The incidence (67,100) and mortality (23,380) of colorectal cancer are lower than that of either breast or lung cancer.

Educational objective:
According to 2019 estimates, the most common cancers (excluding basal and squamous cell skin cancer) in women by order of incidence are breast, lung, and colorectal cancer. In terms of mortality, lung cancer claims the most lives, followed by breast and colorectal cancer.

References

- Cancer statistics, 2019.

Biostatistics	Hematology & Oncology	Prevalence and incidence
Subject	System	Topic

Copyright © UWorld. All rights reserved.

A new biomarker has been identified that allows for the early detection of invasive gastric carcinoma. It has a sensitivity of 89% and a specificity of 85% when compared to endoscopy with multiple biopsies. The test is used in 2 groups: a population in the United States, where 5 out of 100,000 people have gastric cancer, and a population in China, where 100 out of 100,000 people have been diagnosed with the disease. Which of the following is the most accurate statement concerning this new test?

A. Negative predictive value of the test is lower in the American population

B. Positive predictive value of the test is higher in the Chinese population

C. Sensitivity of the test is higher in the American population

D. Specificity of the test is higher in the Chinese population

E. The test is not reliable in the American population

Submit

A new biomarker has been identified that allows for the early detection of invasive gastric carcinoma. It has a sensitivity of 89% and a specificity of 85% when compared to endoscopy with multiple biopsies. The test is used in 2 groups: a population in the United States, where 5 out of 100,000 people have gastric cancer, and a population in China, where 100 out of 100,000 people have been diagnosed with the disease. Which of the following is the most accurate statement concerning this new test?

A. Negative predictive value of the test is lower in the American population (4%)

✓ B. Positive predictive value of the test is higher in the Chinese population (83%)

C. Sensitivity of the test is higher in the American population (3%)

D. Specificity of the test is higher in the Chinese population (3%)

E. The test is not reliable in the American population (3%)

Omitted
Correct answer
B

83%
Answered correctly

03 secs
Time Spent

03/05/2020
Last Updated

Explanation

	Positive condition	Negative condition	
Positive test result	TP	FP	PPV = TP/(TP + FP)
Negative test result	FN	TN	NPV = TN/(TN + FN)

	Positive condition	Negative condition	
Positive test result	TP	FP	PPV = TP/(TP + FP)
Negative test result	FN	TN	NPV = TN/(TN + FN)
	Sensitivity = TP/(TP + FN)	Specificity = TN/(TN + FP)	

FN = false negative; **FP** = false positive; **NPV** = negative predictive value; **PPV** = positive predictive value; **TN** = true negative; **TP** = true positive.

The **positive predictive value** (PPV) is defined as the proportion of individuals with positive test results who actually have the disease. This concept can be understood in terms of probability: given that a patient has a positive test result, what is the likelihood that he or she actually has the disease? PPV depends not only on the sensitivity and specificity of a test, but also on the **prevalence** of the disease in the population being tested. The more common the disease in the population, the more likely it is that a patient with a positive test result actually has the disease.

If the disease is relatively common (eg, gastric cancer in China), there is a higher probability that a patient who tests positive actually has the disease (ie, more likely to be a true positive). However, if the disease is relatively uncommon (eg, gastric cancer in the United States), a patient who tests positive will have a lower probability of actually having the disease (ie, more likely to be a false positive).

(Choice A) Like PPV, the negative predictive value (NPV) depends on disease prevalence. However, NPV is inversely associated with the prevalence: NPV increases as the disease prevalence decreases. This is because the probability of a true negative result is higher in a population with low disease prevalence (eg, gastric cancer in the United States).

(Choices C and D) Sensitivity and specificity are intrinsic test parameters that do not depend on the prevalence of disease in the tested population.

(Choice E) Reliability is a measure of the consistency (reproducibility) of a diagnostic test result (ie, if the test is taken multiple times, will the

FN = false negative; **FP** = false positive; **NPV** = negative predictive value; **PPV** = positive predictive value; **TN** = true negative; **TP** = true positive.

The **positive predictive value** (PPV) is defined as the proportion of individuals with positive test results who actually have the disease. This concept can be understood in terms of probability: given that a patient has a positive test result, what is the likelihood that he or she actually has the disease? PPV depends not only on the sensitivity and specificity of a test, but also on the **prevalence** of the disease in the population being tested. The more common the disease in the population, the more likely it is that a patient with a positive test result actually has the disease.

If the disease is relatively common (eg, gastric cancer in China), there is a higher probability that a patient who tests positive actually has the disease (ie, more likely to be a true positive). However, if the disease is relatively uncommon (eg, gastric cancer in the United States), a patient who tests positive will have a lower probability of actually having the disease (ie, more likely to be a false positive).

(Choice A) Like PPV, the negative predictive value (NPV) depends on disease prevalence. However, NPV is inversely associated with the prevalence: NPV increases as the disease prevalence decreases. This is because the probability of a true negative result is higher in a population with low disease prevalence (eg, gastric cancer in the United States).

(Choices C and D) Sensitivity and specificity are intrinsic test parameters that do not depend on the prevalence of disease in the tested population.

(Choice E) Reliability is a measure of the consistency (reproducibility) of a diagnostic test result (ie, if the test is taken multiple times, will the results be the same?). No information is provided to suggest that the gastric cancer biomarker is unreliable.

Educational objective:
Positive and negative predictive values depend on disease prevalence in the tested population. The sensitivity and specificity of a diagnostic test are not affected by disease prevalence.

Biostatistics	Gastrointestinal & Nutrition	Sensitivity, specificity, positive predictive value PPV, negative predictive value NPV
Subject	System	Topic

Copyright © UWorld. All rights reserved.

FN = false negative; FP = false positive; NPV = negative predictive value; PPV = positive predictive value; TN = true negative; TP = true positive.

The **positive predictive v**
concept can be understoo
the disease? PPV depen
tested. The more commo

If the disease is relatively
disease (ie, more likely to
who tests positive will hav

(Choice A) Like PPV, the
prevalence: NPV increase
with low disease prevalen

(Choices C and D) Sens
population.

(Choice E) Reliability is a
results be the same?). No

Educational objective:
Positive and negative pred
are not affected by diseas

Biostatistics
Subject

Gastro
System

Exhibit Display

Common performance measures for diagnostic tests

Parameter	Definition	Calculation
Sensitivity	The probability of a diseased person testing positive	Sensitivity = True positives / (True positives + False negatives)
Specificity	The probability of a nondiseased person testing negative	Specificity = True negatives / (True negatives + False positives)
Positive predictive value	The probability that disease is present given a positive result	PPV = True positives / (True positives + False positives)
Negative predictive value	The probability that disease is absent given a negative result	NPV = True negatives / (True negatives + False negatives)
Positive likelihood ratio	A ratio representing the likelihood of having the disease given a positive result	LR+ = Sensitivity / (1 − Specificity)
Negative likelihood ratio	A ratio representing the likelihood of having the disease given a negative result	LR− = (1 − Sensitivity) / Specificity

Add To New Card | Existing Card

Copyright © UWorld. All rights reserved.